

A Theory of Digital Library Metadata: *The emergence of Enriching and Filtering*

The thesis is submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Creative Technologies of the University of Portsmouth.

Getaneh Alemu

Supervisors: Dr Brett Stevens, Dr Penny Ross and Jane Chandler

March 2014

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Signed:

Getaneh Alemu

Acknowledgements

First and foremost, I praise the Lord Jesus Christ for His spiritual guidance, love, protection and providence. I am indebted to several individuals who very graciously contributed their insights, critique, support and encouragement to the successful completion of this research. Dr Brett Stevens' intellectual wit, attention to detail and love for rigour greatly contributed to this research project. Equally significant is the guidance and supervision support I received from my second and third supervisors, Dr Penny Ross and Jane Chandler. There was no part of this thesis that escaped their critique and encouraging comments.

I am very much grateful for the financial support I received from the Faculty of Creative and Cultural Industries (CCI) of the University of Portsmouth to cover both my tuition and living expenses. I give special thanks to Dr Simon Claridge (former dean of CCI) and Dr Steve Hand, head of the School of Creative Technologies for their financial support. I also would like to thank Professor David Anderson and Dr Janet Delve for offering me guidance and intellectual support whilst I worked with them.

This thesis benefited from comments and critiques received during various international conferences and seminars including QQML-2011 (Athens), JCDL-2011 (Ottawa), QQML-2012 (Limerick), and IFLA-2012 (Helsinki). Comments received during PhD seminars held in the School of Creative Technologies (Portsmouth) were also helpful. Thus, I thank all who contributed. I would like to thank my former professors of the programme for the International Masters in Digital Library Learning (DILL) of Oslo University College, Parma University and Tallinn University for their encouragement and also facilitation of travel for data collection and suggestions of names of research participants in the areas of metadata and digital libraries. I especially would like to thank Dr Anna Maria Tamaro, Dr Sirje Virkus, Dr Aira Lepik, Dr Sue Myburgh, Katherine Howard and Juan D. Machin Mastromatteo for their professional collaboration.

I extend my heartfelt love for my wife, Bethelhem Getachew Bulfeto. I thank God for bestowing me with such a wonderful soul mate. In my native language Amharic, I would like to say to her: ቤትየ በጣም አወድኻለሁ።

I wish to thank my extended family, especially, my father Ato Agegn Alemu and my mother Woizero Anguach Kassie, for the love and support throughout my life. They are extraordinary parents who sacrificed a lot to make sure I had an education. I wish them long life. I extend my thanks to my in-laws Ato Getachew Bulfeto and Woizero Molash Worku for their love and encouragement. Ato Getachew Bulfeto's thorough proof-reading contributed to the readability of this thesis. Last but not least, I want to thank my friends Dr Mulugeta Mengist, Mekdes Fisseha, Engida Amare, Zelalem Wudineh, Pastor Eyob Negash and everyone at Portsmouth Rhema Church.

Abstract

The ever increasing volume and diversity of information objects, technological advances and rising user expectations is causing libraries to face challenges in adequately describing information objects so as to improve the findability and discoverability of these objects by potential end users. Taking these present metadata challenges into account, this thesis inductively explores and develops overarching concepts and principles that are pertinent within both current standards-based and emerging metadata approaches.

Adopting a Constructivist Grounded Theory Method, this thesis conducted in-depth interviews with 57 purposefully selected participants, comprised of practising librarians, researchers, metadata consultants and library users. The interview data was analysed using three stages of iterative data analysis: open coding, focused coding and theoretical coding. The analysis resulted in the emergence of four Core Categories, namely, metadata **Enriching**, **Linking**, **Openness** and **Filtering**. Further integration of the Core Categories resulted in the emergence of a theory of digital library metadata; **The Theory of Metadata Enriching and Filtering**.

The theory stipulates that metadata that has been enriched, by melding standards-based (a priori) and socially-constructed (post-hoc) metadata, cannot be optimally utilised unless the resulting metadata is contextually and semantically linked to both internal and external information sources. Moreover, in order to exploit the full benefits of such linking, metadata must be made openly accessible, where it can be shared, re-used, mixed and matched, thus reducing metadata duplication. Ultimately, metadata that has been enriched (by linking and being made openly accessible) should be filtered for each user, via a flexible, personalised, and re-configurable interface.

The theory provides a holistic framework demonstrating the interdependence between expert curated and socially-constructed metadata, wherein the former helps to structure the latter, whilst the latter provides diversity to the former. This theory also suggests a conceptual shift from the current metadata principle of sufficiency and necessity, which has resulted in metadata simplicity, to the principle of metadata enriching where information objects are described using a multiplicity of users' perspectives (interpretations). Central to this theory is the consideration of users as pro-active metadata creators rather than mere consumers, whilst librarians are creators of a priori metadata and experts at providing structure, granularity, and interoperability to post-hoc metadata. The theory elegantly delineates metadata functions into two: enriching (metadata content) and filtering (interface). By providing underlying principles, this theory should enable standards-agencies, librarians, and systems developers to better address the changing needs of users as well as to adapt themselves to recent technological advances.

Keywords: *Libraries, Metadata, Metadata Enriching, Metadata Linking, Metadata Openness, Metadata Filtering, Cataloguing, Digital Libraries, Web 2.0, Semantic Web, OPAC, Socially-Constructed Metadata, Standards-Based Metadata, Constructivist Grounded Theory Method*

Table of Contents

Declaration.....	i
Acknowledgements.....	ii
Abstract.....	iii
Table of Contents.....	iv
List of Figures.....	xii
List of Tables	xiv
Abbreviations.....	xv
Dissemination	xvii
Chapter One: Introduction and Research Aims	1
1.1. Introduction.....	1
1.2. Research Aims	3
1.3. Research Scope	3
1.4. Contributions to Knowledge	4
1.5. Structure of Thesis	5
1.6. Conclusion	7
Chapter Two: Standards-based Metadata Approaches.....	8
2.1. Chapter Overview	8
2.2. Definitions.....	8
2.3. Benefits of Metadata	9
2.4. Metadata Actors	10
2.5. Principles for Standards-based Metadata Approaches	10
2.6. Principle of Sufficiency and Necessity	11
2.7. Principle of User Convenience	12
2.8. Principle of Representation.....	13
2.9. Principle of Standardisation	14
2.9.1. The Need for Standardisation	14

2.9.2.	Integration and Interoperability	14
2.9.3.	Guiding Assumptions for Standardisation	15
2.9.4.	Controlled Vocabularies	15
2.9.5.	Metadata Standards for the Semantic Web	17
2.10.	Limitations of Standards-based Metadata Approaches	19
2.11.	Conclusions.....	22
Chapter Three: The Web 2.0 Paradigm and the Emergence of Socially-Constructed Metadata Approaches		
24		
3.1.	Chapter Overview	24
3.2.	Web 2.0 Concepts	24
3.3.	Two-Way Collaboration	24
3.4.	Users as Co-Creators.....	25
3.5.	The Wisdom of Crowds	25
3.6.	Variable Participation	26
3.7.	Openness	26
3.8.	Post-Hoc Quality Control.....	27
3.9.	Web 2.0 Technologies and Applications in Libraries	28
3.10.	The Case of Wikipedia versus Encyclopaedia Britannica	31
3.11.	Limitations of the Web 2.0 Paradigm	34
3.12.	Introducing Socially-Constructed Metadata Approaches	35
3.13.	Conclusions.....	35
Chapter Four: Grounded Theory Methodology		
38		
4.1.	Chapter Overview	38
4.2.	Adopting a Social Constructivist Philosophical Perspective	38
4.3.	The Grounded Theory Method	40
4.4.	Comparison of Grounded Theory Methodologies	41
4.4.1.	Glaserian Grounded Theory.....	41
4.4.2.	Straussian Grounded Theory.....	41
4.4.3.	Charmaz's Constructivist Grounded Theory	42
4.5.	Rationale for Choice of the Constructivist Grounded Theory Method.....	43

4.6.	The Constructivist Grounded Theory Process	44
4.6.1.	Literature Review.....	44
4.6.2.	Data Collection Techniques	45
4.6.3.	Theoretical Sampling	45
4.6.4.	Data Analysis through Coding	45
4.6.5.	Writing Memos	47
4.6.6.	Theoretical Saturation	47
4.6.7.	The Process of Selection of Research Participants	48
4.6.8.	Criteria for Evaluating a Constructivist Grounded Theory	48
4.7.	Conclusions.....	49
Chapter Five: Research Design using the Constructivist Grounded Theory Method		50
5.1.	Overview of Research Design.....	50
5.2.	The Intensive Interviewing Process	51
5.3.	Ethical Review Approval for the Three Studies	53
5.4.	Use of Computer Assisted Qualitative Data Analysis Software (CAQDAS): NVivo 9	54
5.5.	Data Analysis Method and Workflow Using Three-Stages of Constructivist Grounded Theory Method.....	55
5.6.	Conclusions.....	55
Chapter Six: Study One (Proof-of-Concept Study) Preliminary Exploration of the Perspectives of LIS Professionals on Contemporary Metadata Approaches.....		57
6.1.	Overview	57
6.2.	Library as a Place.....	57
6.3.	Obsolescence of Online Public Access Catalogue (OPAC).....	58
6.4.	Preference for Searching Versus Browsing	59
6.5.	Standards-based Metadata Approaches	60
6.6.	Sluggish Adoption of the Web 2.0 Paradigm for Metadata Creation and Utilisation in Libraries	62
6.7.	Sluggish Adoption of Linked Data in Libraries.....	64
6.8.	Bridging Cultural Differences.....	65
6.9.	Collaboration is Essential but Missing.....	67

6.10.	Standards-based Versus Socially-Constructed Metadata Approaches: A Philosophical Difference	68
6.11.	The Need for Theoretical Foundations	68
6.12.	Conclusions.....	69
Chapter Seven: Study Two - Open and Focused Coding of the Perspectives of LIS Professionals on Standards-based Metadata Approaches		72
7.1.	Overview	72
7.2.	Use of Terminologies for Coding and Analysis using Sample Codes from the Three Stages of Coding	75
7.3.	The Current Status of the Adoption of Standards-based Metadata Approaches in Libraries	76
7.3.1.	Research Narrative of Interviewees' on the Current Status of Metadata Standards	76
7.3.2.	Perspectives of Interviewees on Metadata Creators in Current Metadata Practises	82
7.3.3.	A Priori Metadata.....	84
7.4.	Open and Focused Coding Relating to Standards-based Metadata Approaches	86
7.4.1.	Providing Metadata Structure	87
7.4.2.	Providing Metadata Granularity.....	89
7.4.3.	Providing Metadata Provenance	91
7.4.4.	Ensuring Metadata Quality	93
7.4.5.	Ensuring Metadata Simplicity.....	95
7.4.6.	Supporting Metadata Interoperability	96
7.5.	Summary of Categories of Standards-based Metadata Approaches	98
7.6.	Challenges of Standards-based Metadata Approaches	100
7.6.1.	Structural Issues	100
7.6.2.	Metadata Silos.....	101
7.6.3.	Metadata Duplication.....	101
7.6.4.	Slowness in Accommodating and Adopting the Changing Needs of Users	102
7.6.5.	Lack of Engagement with Users	102
7.6.6.	Summary of the Challenges of Standards-based Metadata Approaches.....	103
7.7.	Conclusions.....	104
Chapter Eight: Study Two - Open and Focused Coding of the Perspectives of LIS Professionals on Socially-Constructed Metadata Approaches		105
8.1.	Chapter Overview	105

8.2.	Post-Hoc Metadata Creation	105
8.3.	The Potential Benefit of Involving Users	107
8.4.	Open and Focused Coding Relating to Socially-Constructed Metadata Approaches	110
8.4.1.	Platform for Proactive Metadata Co-creation	112
8.4.2.	Users as Proactive Metadata Co-creators	122
8.4.3.	Metadata Diversity	123
8.4.4.	Metadata Scalability	124
8.4.5.	Collective Metadata Intelligence	125
8.5.	Motivation for Socially-Constructed Metadata Approaches	132
8.5.1.	Reducing Barriers to Contribution	132
8.5.2.	Simplicity, Interestingness and Fun	133
8.5.3.	Re-findability	134
8.5.4.	Sense of Ownership	134
8.5.5.	Altruism and Reputation	135
8.5.6.	Engaging with Users	136
8.6.	Quality Control	136
8.7.	Challenges to Implementing Socially-Constructed Metadata Approaches	140
8.8.	Conclusions	141
Chapter Nine: The Core Category of Metadata Enriching		143
9.1.	Towards a Mixed Metadata Approach	143
9.2.	The Core Category of Metadata Enriching	148
9.2.1.	Metadata Diversity	148
9.2.2.	Metadata Granularity	149
9.2.3.	Metadata Breadth	150
9.3.	Separation of Metadata Creation and Presentation Processes	150
9.4.	Decreasing Marginal Cost of Metadata Creation and Management	150
9.5.	Conclusions	151
Chapter Ten: The Core Category of Metadata Linking		154
10.1.	Chapter Overview	154
10.2.	Contemporary Library Standards and Metadata Linking	154
10.3.	Metadata Linking Technologies	157

10.4.	Re-conceptualising Library Metadata	160
10.4.1.	Breaking Metadata Values into Granular Metadata Statements	162
10.4.2.	Unique Metadata Identifiers.....	163
10.4.3.	Uniquely Identifying Socially-Constructed Metadata.....	164
10.5.	Benefits of Metadata Linking	165
10.6.	Conclusions.....	167
Chapter Eleven: The Core Category of Metadata Openness		170
11.1.	Chapter Overview	170
11.2.	The Core Category of Metadata Openness	170
11.2.1.	Improving Institutional Transparency and Accountability	170
11.2.2.	Return on Investment	172
11.2.3.	Metadata Sharing and Re-usability	172
11.2.4.	Opening Enriched Metadata.....	174
11.2.5.	Improved User Experiences	174
11.3.	Degrees of Metadata Openness and Metadata Licensing.....	175
11.4.	Conclusions.....	177
Chapter Twelve: The Core Category of Metadata Filtering		179
12.1.	Chapter Overview	179
12.2.	Providing Multiple Metadata Views through Post-Hoc Metadata Filtering	179
12.3.	Interface Simplicity.....	181
12.4.	Personalisation	183
12.5.	User Recommendations	185
12.6.	System Recommendations	185
12.7.	Integrated Access to Electronic Resources	187
12.8.	Facilitating Serendipitous Discovery	188
12.9.	Conclusions.....	190
Chapter Thirteen: Study Three - Analysis of the Perspectives of Library Users on Library Resource Usage and the Role of Socially-Constructed Metadata Approaches		192
13.1.	Chapter Overview	192
13.2.	Research Narrative of Interviewees' on the Current Status of Library Systems.....	193

13.2.1.	The Library as a Physical and Virtual Space	193
13.2.2.	The Current Status of Library OPAC.....	195
13.2.3.	The Web as First Stop for Information	197
13.2.3.1.	Convenience of Access	198
13.2.3.2.	Instantaneous Access	198
13.2.3.3.	Currency of Information	199
13.2.3.4.	Access to Full-text	200
13.2.3.5.	About-ness Metadata and Searching inside Books	200
13.3.	Saturating the Core Category of Metadata Filtering	201
13.3.1.	Traversal Across Information Sources.....	202
13.3.2.	Sifting through Relevant Information	204
13.3.3.	Triangulating Information from Disparate Sources	205
13.3.4.	Socially-Constructed Metadata	206
13.3.4.1.	Tagging	207
13.3.4.2.	User Reviews	208
13.3.4.3.	Ratings	209
13.3.4.4.	User Recommendations	209
13.3.4.5.	Book Covers as Metadata	210
13.4.	Conclusions.....	210
Chapter Fourteen: The Emergence of a Theory of Metadata Enriching and Filtering		213
14.1.	Chapter Overview	213
14.2.	Integrating the Four Principles.....	213
14.3.	The Theory of Metadata Enriching and Filtering	214
14.3.1.	Separation of Metadata Content (Enriching) and Interface (Filtering)	215
14.3.2.	Separation of About-ness from Medium.....	215
14.3.3.	Enriching and Filtering as a Non-Deterministic Process	216
14.3.4.	From User-Centred to User-Driven Metadata Enriching and Filtering	216
14.3.5.	Enriching as a Continuous Process	216
14.3.6.	Metadata Diversity better Conforming to Users' Needs	217
14.3.7.	Ubiquitous Linking	217
14.3.8.	'Useful' rather than 'Perfect' Metadata	218

14.3.9. Post-Hoc Filtering	218
14.4. Conclusion	219
Chapter Fifteen: Conclusions	221
15.1. Chapter Overview	221
15.2. Contributions to Theory	221
15.2.1. An Emergent Theory of Metadata Enriching and Filtering	221
15.2.2. Re-Consideration of Librarian and User Roles	222
15.2.3. Re-Conceptualisation of Metadata Quality	223
15.3. Contributions to Practice	223
15.4. Methodological Contribution	226
15.5. Limitations of this Research	227
15.6. Future Research	227
15.7. Concluding Remarks	228
References	229
Appendices	237
Appendix-1: Ethical Approval Letters	237
Appendix-1.1: Ethical Approval of Study One (Proof-of-Concept Study)	237
Appendix-1.2: Ethical Approval Letter for Study Two (Study with LIS Professionals: Researchers, Lecturers and Metadata Consultants)	238
Appendix-1.3: Ethical Approval Letter for Study Three (Study with Library Users)	239
Appendix-2: Interview Questions	240
Appendix-2.1: Preliminary Study on Perspectives of LIS Academics and Post Graduates on Standards-based and Socially-Constructed Metadata Approaches (Study One)	240
Appendix-2.2: A Matrix of Semi-Structured Interview Questions and Constructs (Perspectives of LIS Professionals on Standards-based and Socially-Constructed Metadata Approaches) (Study Two)	242
Appendix-2.3: Semi-structured Interview with Library Users (Study Three)	243
Appendix-3: Glossary	246
Appendix-4: FORM UPR16 (Research Ethics Review Checklist)	248

List of Figures

Figure 1.1: Metadata Categories (based on: Hurley et al., 1999)	3
Figure 5.1: Research Design for Study Two and Study Three (based on Charmaz, 2006)	53
Figure 5.2: NVivo 9 Interface	54
Figure 5.3: Three Stages of Constructivist Grounded Theory Coding (based on Charmaz, 2006)	55
Figure 7.1: Sample Interview Transcript in NVivo 9	73
Figure 7.2: Open Coding Stage Using NVivo 9	74
Figure 7.3: Sample Open Codes of the Open Coding Stage	74
Figure 7.4: Example of use of Open Codes	75
Figure 7.5: Sample Focused Code (Category)	75
Figure 7.6: Core Categories that Emerged from the Theoretical Coding Stage	76
Figure 7.7: Potential Metadata Actors	82
Figure 7.8: Metadata Experts as Major Actors Creating/Curating Basic Metadata	83
Figure 7.9: Data Analysis Showing the Emergence of the Concept of “A Priori Metadata”	85
Figure 7.10: Standards-based Metadata Approaches Instantiating A Priori Metadata	86
Figure 7.11: Open Codes relating to Metadata Structure	87
Figure 7.12: The Category of “Providing Metadata Structure” during Focused Coding	88
Figure 7.13: Open Codes relating to Metadata Granularity	89
Figure 7.14: The Category of Metadata Granularity	91
Figure 7.15: Open Codes relating to Metadata Provenance	92
Figure 7.16: The Category of Metadata Provenance	93
Figure 7.17: Open Codes relating to Metadata Quality	94
Figure 7.18: The Category of Metadata Quality	95
Figure 7.19: Open Codes relating to Metadata Simplicity	96
Figure 7.20: Category of Metadata Simplicity	96
Figure 7.21: Category of Metadata Interoperability	98
Figure 7.22: Summary of Categories of Standards-based Metadata Approaches	99

Figure 7.23: Summary of the Challenges of Standards-based Metadata Approaches	104
Figure 8.1: The Category of Post-Hoc Metadata Creation	106
Figure 8.2: Socially-Constructed Metadata Approach Instantiating Post-Hoc Metadata	106
Figure 8.3: Categories of Socially-Constructed Metadata Approaches	111
Figure 8.4: Components of a Tagging System.....	113
Figure 8.5: Tagging (Open Code during Open Coding)	114
Figure 8.6: User Reviews (Open Code during Open Coding)	115
Figure 8.7: Ratings (Open Code during Open Coding)	116
Figure 8.8: Recommendation Systems (Open Code during Open Coding)	117
Figure 8.9: Crowd Sourcing (Open Code during Open Coding)	119
Figure 8.10: Open Codes relating to Platform for Proactive Metadata Co-creation.....	121
Figure 8. 11: Open Codes relating to Users as Proactive Metadata Co-creators	122
Figure 8.12: The Category of Users as Proactive Metadata Co-Creators	123
Figure 8.13: The Category of Metadata Diversity	124
Figure 8.14: The Category of Metadata Scalability	125
Figure 8.15: The Category of Collective Metadata Intelligence.....	126
Figure 8.16: Metadata Aggregation	128
Figure 8.17: Challenges of Using Socially-Constructed Metadata Approaches.....	140
Figure 8.18: Relationship between Categories of Socially-Constructed Metadata Approaches.....	141
Figure 9.1: Memo on the Complementarity of Standards-based versus Socially-Constructed Metadata Approaches	143
Figure 9.2: Dimensions of Metadata Enriching	148
Figure 9.3: Towards a Mixed Metadata Approach	152
Figure 10.1: Metadata that is not Linked	155
Figure 10.2: Linked Data Technologies.....	159
Figure 10.3: Concepts related to Metadata Linking.....	160
Figure 10.4: Memo on Metadata Linking Technologies.....	161
Figure 10.5: Metadata that is Linked	162

Figure 10.6: Example of Socially-Constructed Metadata (Source LibraryThing.com)	164
Figure 10.7: Example of Book Covers used in Navigation	165
Figure 10.8: The Core Category of Metadata Linking.....	168
Figure 11.1: Degree of Metadata Openness.....	176
Figure 11.2: The Core Category of Metadata Openness.....	177
Figure 12.1: Providing Multiple Views	180
Figure 12.2: Interface Simplicity	182
Figure 12.3: Personalisation.....	184
Figure 12.4: Using User Profiles for Recommendations	186
Figure 12.5: User and System Recommendations	186
Figure 12.6: Integrated Access.....	188
Figure 12.7: Serendipitous Discovery.....	189
Figure 12.8: The Core Category of Metadata Filtering.....	191
Figure 13.1: Saturating Metadata Filtering	202
Figure 13.2: Using Socially-Constructed Metadata.....	206
Figure 13.3: The Core Category of Metadata Filtering.....	211
Figure 14.1: Principles of Metadata Enriching, Linking, Openness and Filtering.....	214
Figure 14.2: Relationship between the Principles of Enriching and Filtering	220
Figure 15.1: The Theory of Metadata Enriching and Filtering	222
Figure 15.2: Changing Librarian and User Roles	223
Figure 15.3: Practical Implications of the Theory of Metadata Enriching and Filtering	226

List of Tables

Table 5.1: Categories of Interviewees	52
Table 13.1: Interviewees for Study Three.....	192

Abbreviations

AACR	Anglo-American Cataloguing Rules
API	Application Programming Interface
BNB	British National Bibliography
CAQDAS	Computer Assisted Qualitative Data Analysis Software
CC:DA	Cataloguing: Description and Access
CC0	Creative Commons Zero
CDWA	Categories for the Description of Works of Art
COTS	Commercial Off-The-Shelf
DC	Dublin Core
DCMI	Dublin Core Metadata Initiative
DDC	Dewey Decimal Classification
DILL	Digital Library Learning
DTF	Date Time Format
EAD	Encoded Archival Description
FRBR	Functional Requirements for Bibliographic Records
HTTP	Hypertext Transfer Protocol
IEEE	Institute of Electrical and Electronics Engineers
IFLA	International Federation of Library Associations
ISBD	International Standard Bibliographic Description
ISBN	International Standard Book Number
ISO	International Organisation for Standardisation
JSC	Joint Steering Committee
LC	Library of Congress
LCSH	Library of Congress Subject Headings
LIS	Library and Information Science
LMS	Library Management Systems
LOM	Learning Object Metadata
MARC	MAchine-Readable Cataloguing
MEDLARS	Medical Literature Analysis and Retrieval System
MeSH	Medical Subject Headings
METS	Metadata Encoding and Transmission Standard

MODS	Metadata Object Description Schema
NISO	National Information Standards Organisation
OCLC	Online Computer Library Centre
OCR	Optical Character Recognition
ONIX	ONline Information eXchange
OPAC	Online Public Access Catalogue
OWL	Web Ontology Language
PDF	Portable Document Format
PREMIS	PREservation Metadata: Implementation Strategies
RDA	Resource Description and Access
RDBMS	Relational Database Management System
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RLG	Research Libraries Group
SPARQL	SPARQL Protocol and RDF Query Language
UNIMARC	UNiversal MACHine-Readable Cataloguing
URI	Uniform Resource Identifier
VIAF	Virtual International Authority File
VRA	Visual Resources Association
W3C	World Wide Web Consortium
WEMI	Work, Expression, Manifestation and Item
XML	Extensible Markup Language

Dissemination

Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012). The Social Space of Metadata: Perspectives of LIS Academics and Postgraduates on Standards-Based and Socially Constructed Metadata Approaches. *Journal of Library Metadata*, 12(4), 311-344. ISSN: 1938-6389.

Alemu, G., Stevens, B., & Ross, P. (2012). Semantic metadata interoperability in digital libraries: a Constructivist Grounded Theory approach. *Bulletin of IEEE Technical Committee on Digital Libraries*, 8 (1). ISSN: 1937-7266.

Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012). Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. *New Library World*, 113(11/22). ISSN: 0307-4803.

Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012). Perspectives of LIS academics and postgraduates on standards-based and socially constructed metadata approaches. *Journal of Qualitative and Quantitative Methods in Libraries*, 1 (1), pp. 99-119. ISSN: 2241-1925.

Alemu, G., Stevens, B., & Ross, P. (2012). Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: A social constructivist approach. *New Library World*, 113(1/2), 38-54. ISSN: 0307-4803.

Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2014). Toward an emerging principle of linking socially-constructed metadata. *Journal of Library Metadata*, 14(2), 103-129. ISSN: 1938-6389.

Conference or Workshop Item

Alemu, G. (2013). *A theoretical framework for the inclusion of socially-constructed metadata in library resource description*. In: BCS Doctoral Consortium 2013, 16 May 2013, BCS London Conference Centre, Covent Garden, London, England.

Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012). *Linked data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data models*. In: World Library and Information Congress 78th IFLA General Conference and Assembly, 11-17 August 2012, Helsinki, Finland.

Alemu, G., Stevens, B., & Ross, P. (2012). *Perspectives of LIS academics and post graduates on standards-based and socially-constructed metadata approaches*. In: 4th International Conference on Qualitative and Quantitative Methods in Libraries, May 22-25, 2012, Limerick, Ireland.

Alemu, G., Stevens, B., & Ross, P. (2011). *A Constructivist Grounded Theory approach to semantic metadata interoperability in digital libraries: preliminary reflections*. In: QQML 2011, 24-27 May 2011, Athens, Greece.

Chapter One: Introduction and Research Aims

1.1. Introduction

Metadata principles which currently underpin the creation and utilisation of metadata functions in libraries have had an enduring history (Denton, 2007; Dunsire, 2009; IFLA, 2009; Lubetzky, 1953; Svenonius, 2000; Wright, 2007). These principles also guided the setup of metadata and cataloguing standards including the Paris Principles in 1961, the Anglo-American Cataloguing Rules (AACR) in 1967, MACHine-Readable Cataloguing (MARC) in the late 1960s, the International Standard Bibliographic Description (ISBD) for Monographic Publications in 1971, Functional Requirements for Bibliographic Records (FRBR) in 1996 and Resource Description and Access (RDA) in 2010 (Denton, 2007; Svenonius, 2000). Thus, the discipline of library and information science has been playing a pivotal role in providing conceptual and technical solutions to the organisation and cataloguing of information objects in libraries (Blair, 2010; IFLA, 2009; Lagoze, 2010; Svenonius, 2000).

However, whilst these metadata principles served libraries relatively well for print collections, with the ever growing volume and diversity of information objects (Blair, 2010; Morville, 2005; Toffler, 1970, 1980; Weinberger, 2007, 2012; Wright, 2007), there arose a need for devising novel approaches for describing, cataloguing and making accessible these collections (Chan & Zeng, 2006; Day, 2003a, 2003b; Duval, Hodgins, Sutton, & Weibel, 2002; Greenberg, 2005; Nilsson, 2010; NISO, 2004). Some of the common metadata standards that aimed at addressing the web environment included the Dublin Core Metadata Standard (DC) in 1995, Metadata Object Description Schema (MODS) in 2002, Metadata Encoding and Transmission Standards (METS) in 2001, and PREservation Metadata: Implementation Strategies (PREMIS) in 2005 (Alemneh, 2009; Alemu, Stevens, & Ross, 2012b; Anderson, Delve, Pinchbeck, & Alemu, 2009; Dekkers, 2002; Gartner, 2008; NISO, 2004; Zeng & Chan, 2006; Zeng & Qin, 2008).

However, critics contend that contemporary metadata approaches have retained some of the constraints inherent in the physical library and card catalogue systems and, hence, fail to rise to the challenge of the present day digital information landscape (Alemu, Stevens, et al., 2012b; Alemu, Stevens, Ross, & Chandler, 2012; Coyle, 2010; Coyle & Hillmann, 2007; Lagoze, 2010; Mathes, 2004; Shirky, 2005; Veltman, 2001; Weinberger, 2005, 2007). As a consequence current metadata principles are criticised for failing to take into account the diversity of cultural, linguistic and local perspectives that abound in library users (Veltman, 2001; Shirky, 2005; Weinberger, 2007).

One of the reasons is that librarians (metadata experts) might not have the requisite expertise in specialised domains in order to adequately describe the semantic (about-ness/content) aspects of information objects (Bowker & Star, 1999; Lagoze, 2010; Shirky, 2005; Veltman, 2001). Chiefly attributed to the nature of language use, one of the most important limitations of conventional metadata approaches, such as subject headings, is the disparity in the terminologies employed by librarians and other metadata experts versus the search terms used by users, potentially affecting the findability of information objects in libraries (Alemu, Stevens, et al., 2012b; Barbosa, 2008; Buckland, 1999; Lu, Park, & Hu, 2010). As Buckland (1999),

Shirky (2005), Veltman (2001) and Weinberger (2007) argue, human beings are highly unlikely to agree on a singular, authoritative and hierarchical classification of objects. Controlling vocabularies and organising information for users presumes some knowledge of the latter's terminologies and search strategies. However, Weinberger (2010, p.184) notes that these presumptions are likely to be incongruent with the actual search behaviours of users. Weinberger (2012) emphasises that librarians cannot possibly anticipate all the various terminologies and languages used by users. Similarly, Bowker and Star (1999) state that although attempts at providing order to information objects are inevitable, these are deeply influenced by social, cultural and political factors. They argue that there are different ways of sorting and cataloguing information (Bowker & Star, 1999).

Furthermore, as the size of collections in libraries grows, librarians increasingly find it difficult to describe every information object, especially those that are digital (Shirky, 2005). In connection with this, Lagoze and Patzke (2011, p. 375) contend that traditional library standards, which had been purposefully designed to describe traditional information sources, have become anachronistic and fail to scale to fully describe new digital genres of information. Lu, Park and Hu (2010) concur stating that in as much as standards-based and expert-created metadata is very good at providing quality metadata, it fails to scale to an ever-increasing volume of digital information that is becoming the prevalent component of collections in institutions, including libraries. The increase in the size of collections, as Lu, Park and Hu (2010) point out, has significant implications on adequately describing and effectively cataloguing information objects in a manner that supports their findability and accessibility.

In contrast to librarian-defined (standards-based) metadata, the Web 2.0 paradigm (O'Reilly, 2005), provides new opportunities for libraries to explore its implications for metadata creation and utilisation. With the growing importance of the emerging Web 2.0 paradigm, it is argued that its implications and impact on libraries and its users cannot be simply ignored (see also Chapter Three), or as Lagoze (2010, p. 37) advises, "the participatory nature of Web 2.0 should not be dismissed as just a popular phenomenon [or fad]". Despite several criticisms, particularly in relation to its lack of structure as well as an absence of editorial quality, authority and credibility (Britannica Inc., 2006; Keen, 2007), many proponents argue that the Web 2.0 paradigm has potential benefits for libraries (Anderson, 2006; Anderson, 2007; Casey & Savastinuk, 2006; Casey & Savastinuk, 2007; Evans, 2009; Farkas, 2007; Guy, 2006; Kroski, 2005; Kroski, 2008; Lagoze, 2010; Maness, 2006; Mathes, 2004; Miller, 2005; Pressley, 2005; Shirky, 2005, 2008; Smith, 2008; Spiteri, Tarulli, & Graybeal, 2010; Spiteri, 2012; Weinberger, 2005, 2007). Whilst some efforts have been made by libraries to introduce aspects of Web 2.0 services (Casey & Savastinuk, 2006; Maness, 2006; Miller, 2005; Spiteri, 2012), their application in metadata creation and utilisation is still limited (Evans, 2009; Lagoze, 2010; Pressley, 2005). Evans (2009, p. 10) notes that "Web 2.0 is an established term, but its implications are not yet fully realised, especially for librarianship," perhaps because, as Lagoze argues, there exists a conceptual incompatibility between traditional library models and the emergent Web 2.0 approaches (Lagoze, 2010, p. 73), an allusion to the absence of theoretical foundations underpinning the combined use of standards-based metadata with socially-constructed metadata approaches.

1.2. Research Aims

In view of current metadata challenges, especially associated with the scarcity of theoretical underpinnings in Library and Information Science supporting metadata functions (Andersen & Skouvig, 2006; Day, 2010; Floridi, 2000; Lawrence, Kehoe, Rieger, Walters, & Kenney, 2000; Lehmann, 2010), it is of paramount importance that a theory of digital library metadata is developed. Underpinning metadata practises with a substantive theory is also important in the light of the centrality of metadata in library functions (Anderson et al., 2009; Chan & Zeng, 2006; Day, 2003a, 2003b; Duval et al., 2002; Nilsson, 2010). It is expected that the theory will contribute to the development, extension and/or refinement of the underlying assumptions and principles that guide and direct the generation and utilisation of metadata in digital libraries.

Taking into account the growing size of library collections, changing users' needs and emerging technological trends, this thesis aims to investigate the current status of existing metadata approaches, and explore the opportunities and alternative solutions that may arise or develop to address contemporary challenges. The thesis mainly considers two strands of metadata: standards-based (expert-created) and socially-constructed (user-created) metadata. This thesis aims to inductively explore the topic, identify and then develop overarching concepts and principles that are pertinent within both current and emerging metadata approaches.

1.3. Research Scope

From among the three categories of metadata, namely descriptive, structural and administrative (Anderson et al., 2009; Gartner, 2008; Hurley, Price-Wilkin, Proffitt, & Besser, 1999; Lavoie & Gartner, 2005; NLA, 1999), only descriptive metadata will be considered in this thesis (see Figure 1.1).

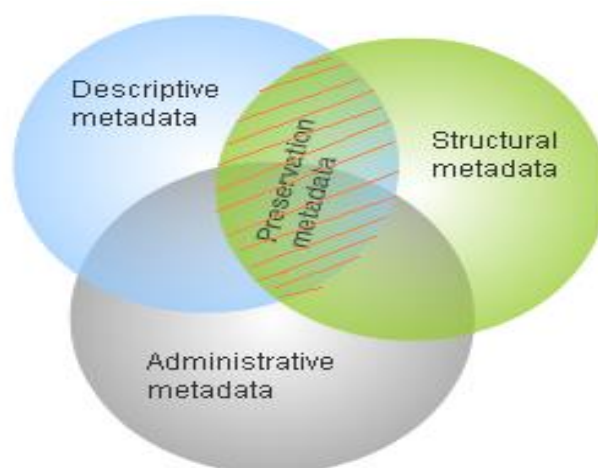


Figure 1.1: Metadata Categories (based on: Hurley et al., 1999)

Whilst there is an overlap among the various metadata categories, descriptive metadata has been chosen as it is primarily concerned with the provision of information about the interpretations (about-ness) of information objects. Thus, this thesis does not concern itself with other categories of metadata, be it administrative or technical. This research does not

attempt to develop a metadata schema, that is to say, the theory will not include explicit specifications for such entities as classes/subclasses, metadata elements (properties) or metadata constraints (data types).

Whilst most of the discussions revolve around digital libraries and electronic access to information, it could also be applied to physical libraries and print information resources. The category of information objects and associated descriptive metadata discussed in this thesis predominately deals with metadata pertaining to books, albeit it also touches on electronic journals and other genres of text information objects. This is because books constitute the major portion of library collections, they are extensively used resources within libraries and books are widely used in every discipline. Thus, any contribution that improves the description of these resources is believed to be of importance. Another compelling reason is the fact that libraries spend quite a significant amount of financial and human resource on describing books. Such investment takes the form of purchases of catalogue records from commercial vendors as well as the acquisition and maintenance of the Online Public Access Catalogues (OPAC); a standardised-metadata database used for searching and locating books and other information sources in libraries. Any improvement that either reduces the outlay for such investments and/or adds consumer value is deemed to be worthwhile and is thus an implicit aim of the suggested theory (1.2). In addition, even if the concepts and principles developed in this thesis have wider implications, the examples and context used cater towards metadata that is pertinent for academic libraries.

1.4. Contributions to Knowledge

In this thesis, using a Constructivist Grounded Theory Method, a rigorous analysis of the perspectives of Library and Information Science professionals, including metadata researchers, academics, consultants, and practising librarians working at world-renowned international, national and academic libraries (see Chapters Seven through Twelve) as well as library users was conducted (see Chapter Thirteen). From the analysis, four overarching metadata Core Categories emerged, namely metadata enriching (Chapter Nine), Metadata linking (Chapter Ten), Metadata openness (Chapter Eleven) and Metadata filtering (Chapter Twelve). The integration of these Core Categories resulted in the emergence of a **Theory of Metadata Enriching and Filtering** (Chapter Fourteen). The theory elegantly captures and presents several overarching complex concepts and principles subsumed under a mixed metadata approach where both standards-based and socially-constructed metadata can be used in combination, each offering what it does best.

The theory suggests a paradigm shift from an objectivistic ontology, wherein a single metadata interpretation (expert-created) exists, to an interpretative (social constructivist) ontological point of view, wherein a multitude of metadata interpretations can co-exist. Methodologically, the paradigm suggests a shift:

- From metadata simplicity to metadata enriching (15.3.2.1);
- From human-readable and non-linked metadata to granularly structured, uniquely identified, machine process-able and interlinked metadata (15.3.2.2);

- From metadata that is locked to metadata openness (15.3.2.3);
- From a single (global) OPAC interface to metadata that can be filtered through re-configurable interfaces (15.3.2.4).

Practically, the emergent theory, empirically grounded in the data collected from the perspectives of LIS researchers, librarians, metadata experts and library users, has major implications (Chapter Fifteen) for standards-setting agencies (15.3.1), libraries (15.3.2), library management system developers (15.3.3) and library users (15.3.4). These wider implications imply the emergence of a new metadata paradigm.

The theory of Metadata Enriching and Filtering is thus considered an original contribution to the development, extension and refinement of contemporary metadata assumptions and principles. The theory provides a holistic framework demonstrating the interdependence between expert curated and socially-constructed metadata, wherein the former helps to structure the latter, whilst the latter provides diversity to the former. The emergent theory of digital library metadata, although buttressed by two seemingly simple concepts, is nonetheless a first rigorous attempt in LIS research to integrate the concepts of enriching, linking, openness and filtering into principles and subsequently into a holistic digital library metadata theory. By making the case for mixed metadata (standards-based and socially-constructed metadata), the theory brings into focus the importance of re-conceptualising current metadata principles, mainly suggesting a shift from objectivistic and deterministic metadata approaches which chiefly focus on metadata simplicity (2.6) to a social constructivist and non-deterministic (14.3.3) continuous and evolving process of metadata enriching (14.3.5). The theory is therefore an elegant representation of otherwise complex concepts and principles. The elegance of the theory is its ability to capture the complexity in its implications.

Methodologically, the rigorous adoption of the Constructivist Grounded Theory Method (see also Chapters Four and Five) showed that the methodology is a viable approach to explore and investigate emerging research areas in LIS, especially in the areas where diverse views prevail due to its novelty. This method supports the abstraction of a theory from existing practises and subsequently allows new practise to be derived from the theory.

The emergence of the Theory of Metadata Enriching and Filtering contributes to the development of the discipline of Library and Information Science in general and the field of digital libraries and metadata in particular, thus partly addressing the absence of theoretical foundations (Andersen & Skouvig, 2006; Day, 2010; Floridi, 2000; Lehmann, 2010).

1.5. Structure of Thesis

In **Chapter One**, the thesis commences by broadly identifying current metadata challenges, such as the growing size of collections in libraries, an absence of specialist knowledge by librarians to accurately describe collections, disparity in the vocabulary of librarians and users, and the challenges associated with the failure to represent user terminologies. In light of these challenges, the chapter indicates the importance of exploring alternative approaches,

and suggests the need to explore the role of the Web 2.0 paradigm for metadata functions. The chapter also describes the research aims, scope, and contributions to knowledge.

In order to provide context to the research problem, a review of the literature pertaining to the two strands of metadata, i.e. standards-based and socially-constructed metadata approaches is presented in **Chapters Two** and **Three**, respectively. Then **Chapter Four** discusses the choice of an interpretive ontology, a social constructivist epistemology and an inductive methodology. The chapter also justifies the rationale for the choice of Grounded Theory Method and provides an account of the reasons for the choice of the Constructivist Grounded Theory Method. The chapter then reviews the three data analysis stages of the Constructivist Grounded Theory coding.

Chapter Five explains the practical research design and the processes adopted. As this chapter describes, this research is informed by three separate but interrelated studies. The chapter also provides a rationale for the use of Computer Assisted Qualitative Data Analysis Software (CAQDAS).

Chapter Six presents analysis and findings of a preliminary (proof-of-concept) study (Study One). This had been conducted to capture the perspectives of LIS academics and post-graduate students on standards-based and socially-constructed metadata, and thus identify preliminary themes to design the main study.

Chapter Seven presents the findings of Study Two (i.e. interviews with practising librarians, LIS researchers and metadata consultants). Based on the Focused Coding, the Categories discussed in this chapter include metadata structure, granularity, provenance, quality, simplicity and interoperability. The chapter also discusses the current status and challenges of standards-based metadata approaches.

Chapter Eight presents findings in relation to socially-constructed metadata approaches. The Categories emerged from the Focused Coding data analysis are presented. These include a platform for metadata co-creation, users as proactive metadata co-creators, metadata diversity, scalability and collective metadata intelligence.

Based on the emerging Core Categories (Theoretical Coding), **Chapter Nine** discusses the first Core Category, i.e. Metadata Enriching, followed by **Chapters Ten, Eleven** and **Twelve** discussing the Core Categories of metadata linking, metadata openness and metadata filtering, respectively.

Chapter Thirteen presents the processes and results of Study Three, i.e. interviews with library users. By integrating the four Core Categories, **Chapter Fourteen** presents the **Theory of Metadata Enriching and Filtering**. Finally, **Chapter Fifteen** presents the summary of the findings of the thesis and contributions of this to knowledge in library and information science in general and metadata research in particular. The chapter also presents the implications of the **theory of Metadata Enriching and Filtering** on standards-setting-agencies, libraries, library system developers and library users. The final chapter also identifies the limitations of this research and pinpoints future research areas on the topic.

1.6. Conclusion

This chapter introduced the research problem by identifying current metadata issues in relation to the importance of adequately and accurately describing the increasing size of information objects in libraries in a manner that incorporates the diversity of interpretations of users. The chapter suggested that, in view of the Web 2.0 paradigm and changes in users' expectations, the importance of exploring alternative metadata approaches becomes significant. As described in this chapter, the scope of this research is limited to descriptive metadata, but within that it is aimed at developing a theory of digital library metadata.

Chapter Two: Standards-based Metadata Approaches

2.1. Chapter Overview

As discussed in Chapter Four (4.6.1), whilst some grounded theorists such as Glaser (1978) and Strauss and Corbin (1990) argue that review of related literature should wait until the completion of the empirical data analysis, in this thesis, following Charmaz (2006) the review is made at an earlier stage of the study in order to provide context and background for the study, especially in relation to the two strands of metadata. It is perhaps worth disclosing that this researcher's interest in metadata and digital libraries was based on more than ten years of educational and work experiences in the field of Library and Information Science. As the background suggests, one cannot claim avoidance of preconception to extant concepts and theories in the topic under investigation. Even though, the Web 2.0 paradigm and its implication for metadata creation and utilisation, especially in view of developing a theory, is relatively unexplored, the researcher has developed a keen interest in the area of socially-constructed metadata approaches (Alemu, Stevens, & Ross, 2012a; Alemu, Stevens, et al., 2012b). In accordance with this, in this research, the review of related literature was considered an important iterative process which was conducted at quite an early stage but revised and refined as the research progressed.

In light of this, Chapter Two provides a review of relevant literature in relation to standards-based metadata approaches. It provides definitions and functions of metadata and discusses current metadata principles that underpin current metadata practises. Finally, the limitations of current standards-based metadata approaches are identified and possible alternatives are pinpointed.

2.2. Definitions

The term metadata is defined as data about data. However, Lavoie and Gartner (2005) and Day (2005) argue that this definition is unhelpful and they suggest that metadata should be defined in relation to its functions. Such definition is given by the US National Information Standards Organisation (NISO, 2004), which defines metadata as “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource”. By providing descriptive (author, title and subject), administrative (identity, provenance, rights, contextual and technical) and structural information (relations with other information objects), metadata plays an important function in digital libraries to support the findability and discoverability of information objects by users and also librarians.

Metadata is an important component of any library and digital repository system (Anderson et al., 2009; Caplan, 2003; Chan & Zeng, 2006; Day, 2000, 2001, 2003a, 2003b; Duval et al., 2002; Gartner, 2008; Greenberg, 2005; Greenberg, Sutton, & Campbell, 2005; Hurley et al., 1999; Lagoze, 2010; Lavoie, 2004; Nilsson, 2010; NISO, 2004; Zeng & Qin, 2008). In the literature on metadata, the terms schema, scheme and metadata element set are used, interchangeably, to refer to the formal specifications of attributes (characteristics) employed

for representing information resources (Lagoze, 2001a, 2010; NISO, 2004). However, some such as Greenberg (2005) prefer to use the term scheme (instead of schema) when they refer to metadata elements of a metadata standard.

Thus, for clarity and following Lagoze (2010), in this research the term metadata schema is used to refer to the set of metadata elements (also called fields) in a metadata standard. The Committee on Cataloguing: Description and Access (CC:DA) defines a metadata schema as a structured specification of the elements, values and constraints (rules) that are designed in such a way that identification, discovery and utilisation of information is done in a consistent manner (CC:DA, 2000). Metadata schemas are also, commonly, referred to as standards, even though some schemas are de facto, rather than de jure, standards. Some metadata standards-setting agencies use different terminologies to refer to their metadata components. For instance, Dublin Core uses the phrase ‘element set’ (Weibel, 1999; Weibel & Koch, 2000; Zeng & Qin, 2008), whereas PREMIS uses the term ‘semantic unit’ (Alemneh, 2009; Caplan, 2009; Dappert & Enders, 2008; Guenther, 2005; Guenther, 2008; Lavoie & Gartner, 2005; Lavoie, 2008; OCLC/RLG, 2008).

According to St. Pierre and LaPlant (1998) any metadata schema should include a unique name for each metadata element, a human-readable definition, as well as information as to whether a value for a metadata element is mandatory or optional, repeatable or unique. Similarly, Nagamori and Sugimoto (2006) state that metadata schemas should include names for metadata elements, as well as sources from which values can be instantiated (such as one of the following controlled vocabularies: ISO-8601, DCMI Type Vocabulary, LCSH and DDC). In connection with this, Bruce and Hillmann (2004) advise that metadata schemas should contain element sets that are likely to be usable, complete, accurate and accessible.

2.3. Benefits of Metadata

The solution to the information overload problem of the web, as Weinberger (2007, p. 13) notes, is to create yet more information about the information, i.e. metadata. As Zeng and Qin (2008, p. 3) note, metadata is “the invisible hand” that enables effective information organisation. In recognition of this importance, significant investments have been made to specify metadata standards by a number of national, multinational and international initiatives in order to describe collections and to enhance the findability and discoverability of information objects (Alemneh, 2009; Alemu, Stevens, et al., 2012b; Anderson et al., 2009; Baker, 1998, 2000; Caplan, 2009; Dekkers, 2002; Dekkers & Weibel, 2002; Duval et al., 2002; Gartner, 2008; Lagoze, 2001a, 2001b, 2010; Lavoie & Gartner, 2005; Zeng & Chan, 2006; Zeng & Qin, 2008). Metadata is thus considered an essential function of libraries, for example, as Deanna Marcum, reported “the Library of Congress is investing in cataloguing at the rate of forty-four million dollars a year” (Marcum, 2005, p. 1). With regard to this, some of the common metadata standards that aim to address the web environment include the Dublin Core Metadata Standard (DC), Metadata Object Description Schema (MODS), Metadata Encoding and Transmission Standards (METS) and PREservation Metadata: Implementation Strategies (PREMIS) (Alemneh, 2009; Alemu, Stevens, et al., 2012b; Anderson et al., 2009; Dekkers, 2002; Gartner, 2008; NISO, 2004; Zeng & Chan, 2006; Zeng & Qin, 2008).

2.4. Metadata Actors

Broadly speaking, there are four groups of metadata creators: librarians, authors, users and machines (Mathes, 2004). Metadata created by librarians usually follows detailed sets of rules and workflows. To this effect, librarians receive formal trainings in cataloguing/metadata and the use and application of standards and principles, such as AACR, DDC, MARC, LCSH, MeSH, RDA and FRBR, to mention but a few (Lagoze, 2010; Lavoie & Gartner, 2005; Svenonius, 2000; Zeng & Chan, 2006; Zeng & Qin, 2008). Consequently, the metadata created by librarians is considered to be of relatively higher quality, particularly with regard to accuracy, completeness and consistency (Bruce & Hillmann 2004; Haynes, 2004; Park & Tosaka, 2010). However, librarian-generated metadata does not scale well, in view of the growing size of library collections (Alemu, et al., 2012; Barbosa, 2008; Lu, et al., 2010; Mathes, 2004; Shirky, 2005; Smith, 2008; Weinberger, 2007). Furthermore, as Mathes (2004) points out, yet another main limitation of both librarian and author-created metadata is that “the intended and unintended eventual users of information are disconnected from the process”. Moreover, Doctorow (2001) contends that authors may want to see their works appear at the top of search engines listings, and, hence, may fail to accurately represent their contents, thus indicating the limitations of author-created metadata.

In standards-based approaches, values assigned to elements during metadata creation, are rigorously checked for consistency before the metadata is made available to users through the Online Public Access Catalogue (OPAC). These tasks often require rigorous adherence to applicable cataloguing and classification rules and regulations. Thus, metadata generation is chiefly done by experts (librarians), implicitly presuming that metadata creation and management is the sole prerogative of the librarian. Libraries have so far employed cataloguers and classifiers to register their collections and have adopted one or more of several extant cataloguing standards, such as Anglo-American Cataloguing Rules (AACR), DDC and LC. However, as the volume of digital information increases, assigning metadata to tens, if not hundreds, of millions of digital information objects has become a progressively more expensive endeavour (see discussions on metadata actors in Study Two, 7.3.2 & 7.3.3). Finding a viable approach for creating, maintaining and utilising metadata is therefore essential. As a consequence, libraries and other cultural heritage institutions attempted to derive some common metadata and cataloguing principles.

2.5. Principles for Standards-based Metadata Approaches

Foundational principles, rather than case-based rules for library cataloguing, provide professionals with a sound backdrop when constructing coherent catalogues of information objects (Lubetzky, 1955). A principle is one or more sets of abstract, general propositions that guide practitioners in the development of guidelines and rules, which in turn direct practice (Duval et al., 2002; IFLA, 2009; Svenonius, 2000). With the growth in the size of collections as well as diversity in mediums of information, libraries were obliged to develop disparate sets of rules for describing their information objects, whereby “the multiplicity and variety [of these rules] seemed bewildering,” suggesting the need for establishing general principles that guide practice (Lubetzky, 1955, p. 183). Lubetzky (1955) advocates the tenet of constructing library cataloguing standards on theoretically coherent foundational principles

rather than upon case-based rules. Lubetzky (1955) further maintains that the history of cataloguing has been characterised as one that has focused more on rules than on principles. Nevertheless, there have been efforts at developing overarching principles for standards. For example, the following are some of the standards that are based on established principles: AACR, MARC, ISBD, FRBR and RDA (Denton, 2007; Dunsire 2009; Lubetzky, 1955; Svenonius, 2000). According to IFLA (2009) and Svenonius (2000), some of the major foundational principles that underpin standards-based metadata approaches include:

- The principle of sufficiency and necessity
- The principle of user convenience
- The principle of representation
- The principle of standardisation.

2.6. Principle of Sufficiency and Necessity

The origin of the principle of sufficiency and necessity comes from history of modern cataloguing (IFLA, 2009; Hoffman, 2009; OCLC, 2009; Spiteri, 2012; Svenonius, 2000). Pioneers of library cataloguing such as Cutter, Panizzi and Lubetzky are said to have advocated for a metadata approach that caters for simplicity (Svenonius, 2000). Building on earlier works, the International Federation of Library Associations (IFLA, 2009) also used this principle to underpin the design of its Functional Requirements for Bibliographic Records (FRBR) model, the most dominant model that serves as a reference when designing future metadata and discovery systems.

As described by IFLA (2009, p.2), the principle of sufficiency and necessity states that “only those data elements in descriptions and controlled forms of names for access that are required to fulfil user tasks and are essential to uniquely identify an entity should be included”. It is tacitly assumed that superfluous metadata distracts and confuses users, thus the principle states that metadata should be kept minimal and simple by recording only the metadata pertinent to find information objects. This principle was in accord with IFLA’s principle of economy, where it states that “when alternative ways exist to achieve a goal, preference should be given to the way that best furthers overall economy (i.e. the least cost or the simplest approach)” (IFLA, 2009, p.2). Since creating metadata records is an expensive endeavour, the principle of sufficiency and necessity is considered to provide efficiency, and, thus, is considered a cost saving mechanism. The principle assumes that the cost of metadata increases proportionally with the number of metadata elements considered, thus metadata agencies are encouraged to include metadata elements that are only considered essential (Svenonius, 2000, p.76). As Svenonius points out, standards-setting agencies (e.g. IFLA), bibliographic data suppliers (e.g. OCLC) and library metadata experts largely predicated their efforts on the principle of sufficiency and necessity. The principle suggests that librarians record metadata that is only of significant value, subsequently, as Svenonius (2000, p.77) notes, librarians apply the rule of Occam’s razor, whereby metadata that is considered superfluous is eliminated, resulting therefore in metadata simplicity.

2.7. Principle of User Convenience

The principle of user convenience addresses the importance of designing metadata systems with users in mind (Hoffman, 2009; IFLA, 2009; Spiteri, 2012; Svenonius, 2000). Ranganathan, as early as 1931, stressed the centrality of the user in his publication "The Five Laws of Library Science" (Denton, 2007, p. 44). Ranganathan's five "laws" were: books are for use; every reader its book; every book its reader; save the time of the reader; and library is a growing organism (Welsh & Batley, 2012). One of the first attempts at putting this principle into practice was the move to giving users access to the library catalogue (Denton, 2007; IFLA, 2009; Svenonius, 2000). Up until the publication of "Rules for the Compilation of the Catalogue," by Sir Anthony Panizzi in 1841, the main users of library catalogues had been librarians themselves (Wright, 2007). However, the growth in size of library collections had made it imperative that the catalogue itself be made open to the public (Wright, 2007). The main objective of Panizzi's rules was to provide access to the right book (including its various editions, if any, as well as to others on a related topic) to the right user, at the right time (Lubetzky, 1955; Svenonius, 2000; Welsh & Batley, 2012). Panizzi successfully pursued the practice of opening up the library catalogue for consultation by users (Wright, 2007, p. 169). The practice of opening up the catalogue to users is part and parcel of the general principle of user-convenience, wherein it is stipulated that metadata decisions including the choice of controlled vocabularies reflect users' needs (IFLA, 2009; Spiteri, 2012; Svenonius, 2000). According to Svenonius (2000), important concerns in connection with this include taking users' preferences into account, based on their profiles.

User-convenience necessarily implies that metadata in standards-based systems reflect the metadata needs of each user, including the terminologies used for describing information objects, the management of metadata and the display of OPAC results. In other words, a metadata system that fully complies with the principle of user-convenience must incorporate practical provisions for delivering such personalised and user-centred services (Hoffman, 2009). Customisation may consist of changing the way the metadata is displayed, filtered, or navigated through. However, it might be difficult to fully meet each and every one of these personalised and idiosyncratic needs in physical libraries.

Standards-based metadata approaches, although acknowledging the importance of the principle of user-convenience, face serious challenges when attempting to adhere to the principle as a whole (Hoffman, 2009). Some of the constraints can be attributed to the inherent limitations of physical libraries and the costs associated with metadata creation and maintenance. Hoffman (2009) maintains that, although standards-setting agencies and libraries are prone to proclaiming that their systems are user-focused/user-centred, contemporary standards, including such prominent ones as FRBR and RDA, tend to focus on devising universal and uniform systems. According to the author, one common illustration of uniformity in metadata standards is the way metadata is displayed in OPACs, which mostly consist of single, uniform interfaces and homogenous navigation/search techniques, a consequence of the objectivistic focus of standards-setting agencies as well as implementing institutions, such as libraries (Hoffman, 2009). Other examples include hierarchies and categories, which are fixed and, hence, non-customisable by either librarians or users (Shirky,

2005). Hoffman acknowledges that customisation has not been made possible due to the fact that most bibliographic data and metadata systems are acquired from external sources and, thus, are often difficult to modify. In addition, Hoffman (2009) points to an absence of mechanisms for maintaining user profiles, which could have been harnessed for customisation and personalisation.

Perhaps the most important constraint for realising the principle of user-convenience arises from a basic assumption adopted by standards-based approaches, namely, the treatment of the user as a passive consumer. As noted by Hoffman (2009, p.633), “the principle of user convenience assumes that cataloguers can objectively determine users’ needs and will know how to customise bibliographic records to meet these needs”. This attempt, at a supposedly, objective speculation on user’s preferences, results in putting all users “into one basket”, thereby failing to address the heterogeneity that exists amongst them (Hoffman, 2009). However, as long as the principle of user-convenience is hinged upon the assumption that users are passive consumers of metadata and that metadata is solely created by librarians, complying with the principle is bound to be difficult.

2.8. Principle of Representation

According to Svenonius (2000, p. 71) the principle of representation stipulates that metadata should objectively and accurately represent an information object, and that such representations avoid idiosyncratic description. According to IFLA (2009), “descriptions and controlled forms of names should be based on the way that an entity describes itself,” which implies that the metadata should correspond to information that is contained on the information object itself, for instance in the title page of a book. Elaborating on this principle, Svenonius (2000, p. 71) states that: “truthfully transcribing how a document represents itself is necessary for the identification and communication of bibliographic information. The metadata in a description are truthfully transcribed, or warranted, if (1) they come from a specified source and; (2) they are copied in the form in which they appear there, except for capitalisation and punctuation. These two conditions ensure the likelihood that two people describing the same bibliographic entity will create identical descriptions.”

The principle of representation targets such goals as the reduction of metadata creation costs as well as the provisioning of accurate descriptions. As is the case for the principle of user-convenience, it presumes librarians as sole creators of metadata. Furthermore, the principle compels librarians to use authoritative sources of metadata, automatically ruling out the use of vocabularies that might be considered informal or non-academic. It states that, “given a choice of several sources, data are to be taken from the source that gives the fullest, clearest and most authoritative information” (Svenonius, 2000). Moreover, the principle prohibits inclusion of any two inconsistent descriptions of an information object, thereby precluding the co-existence of diverse, potentially conflicting, interpretations of information objects. This is, most probably, a consequence of the objectivist tendency, an approach that is incapable of accommodating the diversity of cultural, linguistic and personal perspectives that library users can have (Shirky, 2005; Veltman, 2001; Weinberger, 2007). Finally, according to this principle, the title page is the main source of information for authoritatively

creating a representative description (Svenonius, 2000). The problem with this stricture, however, is the fact that, should the sole source of data be limited to the title page (or even other parts of the information object), metadata that comes from other sources, including user-generated metadata (see 3.12), is precluded (Svenonius, 2000).

2.9. Principle of Standardisation

With the aim of facilitating metadata consistency and interoperability, current metadata approaches are characterised by standardisation (IFLA, 2009; NISO, 2007).

2.9.1. The Need for Standardisation

According to IFLA (2009, p. 2), in the context of cataloguing, the principle of standardisation states that, “descriptions and construction of access points should be standardised as far as possible”. It is indicated that standardisation leads to consistency, which in turn facilitates metadata sharing of bibliographic records (IFLA, 2009). The National Information Standards Organisation (NISO, 2007) also acknowledges standardisation as a core principle. It is evident that such standardisation minimizes costs, ensures consistency, and facilitates the exchange of metadata across digital platforms. Conformity to established standards is also considered a prerequisite for achieving interoperability. Examples of established standards include MARC, MODS, DC, EAD, METS, CDWA Lite and VRA Core (NISO, 2007).

Even though the need for standardisation has been recognised in various domains, including digital libraries, it is important to note that there are a host of problems when implementing them. One of these problems arises as a consequence of the existence of multiple standards that accomplish the same function. Furthermore, institutions may, in reality, prefer to use their own local metadata schemas, rather than adopting international standards, for several reasons, including politics, competition, or an institutional need for innovation and experimentation (Alemu, Stevens, et al., 2012b; Veltman, 2001). As a result, institutions tend to use either local standards (Veltman, 2001) or a diverse mix of local, regional and international ones (Coyle, 2009; Gartner, 2008; Heery & Patel, 2000).

2.9.2. Integration and Interoperability

The objectives of metadata standardisation include the support of metadata integration and interoperability (Alemu, Stevens, et al., 2012b; Day, 2000; Gartner, 2008; Haslhofer & Klas, 2010). The issues of metadata interoperability are variously discussed in the metadata literature (Alemu, Stevens, et al., 2012b; Arms et al., 2002; Bailer, 2007; Bearman, Miller, Rust, Trant, & Weibel, 1999; Bianchi & Petrone, 2001; Cabinet Office, 2004; Chan & Zeng, 2006; Day, 2000, 2003a, 2003b; Dekkers, 2002; Dempsey, 2004; EC, 2004, 2010; Gartner, 2008; Gill & Miller, 2002; Guenther, 2008; Haslhofer & Klas, 2010; Miller, 2000; Nagamori & Sugimoto, 2006; Nilsson, 2010; Nilsson, Johnston, Naeve, & Powell, 2006; Ouksel & Sheth, 1999; Rothenberg, 2008; Veltman, 2001; Zeng & Chan, 2006). The existence of several ‘international’ metadata standards, coupled with the proliferation of several “in-house” metadata schemas, as a review of the metadata literature shows, has exacerbated the problems of metadata interoperability (Alemu, Stevens, et al., 2012b; Day, 2000; Haslhofer &

Klas, 2010; Veltman, 2001). Under these circumstances, achieving metadata interoperability, with the adoption of a single standard, becomes a daunting task (Alemu, Stevens, et al., 2012b; Day, 2000; Haslhofer & Klas, 2010; Veltman, 2001). In situations where several metadata standards co-exist, some of the approaches that have been employed to effect metadata interoperability include the use of metadata derivation, application profiles, metadata-cross walks (metadata matching), metadata registries and the use of semantic web technologies (Chan and Zeng, 2006; Day, 2003b; Nagamori and Sugimoto, 2006; NISO, 2004). However, it has been amply demonstrated that even the whole scale adoption of all these approaches and methods cannot provide the required semantic interoperability for effective cross-searching, content sharing and information integration (Nilsson, 2010). Hence, metadata interoperability still remains a big challenge. However, semantic metadata interoperability, whilst an important concern for future research, is beyond the scope of this thesis.

2.9.3. Guiding Assumptions for Standardisation

In relation to metadata standards, Duval, et al. (2002) also identify modularity, extensibility, refinement and multilingualism as guiding principles of any metadata schema. Stressing the benefits of a modular approach, Duval, et al. (2002) for example argues that “in a modular metadata world, data elements from different schemas as well as vocabularies and other building blocks can be combined in a syntactically and semantically interoperable way.” As Duval, et al. (2002) explains from such modularity, designers can benefit from reusability. The refinement feature of metadata schemas is exemplified by the Dublin Core. The 15 metadata elements can be refined using qualifiers such as *Illustrator*, *Author*, *Composer*, or *Sculptor* for the general term *Creator*. And secondly, using different encoding schemes such as the W3C date and time format (W3C-DTF), the values can be constrained into a more stringent requirement so as to maintain consistency and accuracy. Using the W3C-DTF scheme, the date January 5th, 2012 is represented as 2012-01-05 in a YYYY-MM-DD format (Duval et al., 2002). The multilingualism principle indicates the ability of the schema to represent different languages and reflects multiculturalism by, for example, allowing different characters, calendars, and direction of text displayed and read, naming order and related nitty-gritty of languages (Duval et al., 2002).

2.9.4. Controlled Vocabularies

Controlled vocabularies, which include such instantiations as taxonomies, thesauri and ontologies, are considered in the metadata literature as mechanisms to enforce standardisation in metadata values. A taxonomy is a controlled list of vocabulary or terms, with a hierarchical structure of broader/narrower or parent/child relationships amongst them (Barbosa, 2008; Hedden, 2010). Similarly, Garshol (2004) describes a taxonomy as a “subject-based classification that arranges the terms in the controlled vocabulary into a hierarchy” of broader and narrower terms. Hedden (2010, p. 1) provides an extensive review of the role of taxonomies in information organisation and discovery. A thesaurus extends these taxonomies with additional relationships, such as related term(s), synonyms, preferred term and scope

notes (Garshol, 2004). The third type of controlled vocabularies is ontology, which is defined by Gruber (2007) as an explicit specification of a conceptualisation.

In relation to controlled vocabularies, there are certain assumptions and principles that underpin its construction. One such principle is the assumption that “categories are mutually exclusive,” which implies that an object can belong to one, and only one, branch in a hierarchy (Bowker & Star, 1999). Another principle is completeness of the classification scheme itself, whereby, for example, “a botanical classifier would not simply ignore a newly discovered plant, but would always strive to name it”. According to Bowker and Star (1999, p. 10), “a classification [scheme] is a set of boxes (metaphorical or literal) into which things can be put, to then do some kind of work - bureaucratic or knowledge production”.

According to Bowker and Star (1999), although, classification schemes are ubiquitous, the veracity of some of the underlying principles, such as mutual exclusivity and completeness, are debateable. They contend that “each [classification scheme] and each category valorises some point of view and silences another”. The main shortcomings of a classification scheme come to the fore when such a scheme attains the status of a standard. This is almost always inescapable, but, nonetheless such inherent limitations should be taken into account and their ultimate consequences carefully weighed whenever such standards are designed and implemented. Concurring with this, Doctorow (2001) states that classification schemes “are not neutral”. Doctorow singles out human fallibility (lies, errors and biases) as confounding elements for so-called neutral schemes and reliable metadata. Hence, Doctorow contends, explicit, human-generated, metadata should be used with “a large pinch of salt”. Reliable metadata, he argues, can only be obtained from an implicit harnessing of metadata that is left behind by users without them intending to explicitly tag or label information objects. Similarly, Bowker and Star (1999, p. 32) argue that each classification scheme is bias-laden. The bias is often revealed when the language (terminologies) used in these schemes significantly diverges from the one employed by its users.

When librarians create or purchase metadata records, it is the conventional wisdom that these records should comply with one or more of the internationally recognised metadata standards or structures; such as MARC records (Zeng & Qin, 2008). Field contents are derived from recognised controlled vocabularies, such as LCSH, MeSH, Art and Architecture Thesaurus and related other taxonomies and thesauri. In addition, internationally recognised encoding schemes are enforced to ensure consistency in the format of the records (for example, W3C-DTF (ISO 8601) for Date; and RFC-4646 for Language) (Zeng & Qin, 2008). Metadata fields (such as Author, Title, Date of Publication and ISBN) and their corresponding values are purposefully made objective, as, for example, there is no reason to question whether the author of “Animal Farm” is George Orwell; excepting the fact that this is his pen name (his real name being Eric Arthur Blair). Similarly, one doesn’t gain any benefits by recording its publication date differently. However, the challenge comes with regard to the about-ness of the book, i.e. there are possible interpretation differences that are bound to ensue as soon as readers of the book are asked to describe it with keywords, or enter in to conversation about their perceptions regarding its principal characters. However, as explained in the example

above, conventional library metadata practises are based mainly on predetermined schema (structure of metadata elements), internationally accepted encoding schemes and authorised controlled value lists (Alexander, 2006).

It is indicated that stringent metadata standardisation implies strict adherence to and consistent implementation of a single standard by a library (see Chan and Zeng, 2006). As Svenonius (2000, pp. 80-82) points out, some of the limitations of the principle of standardisation include:

- Difficulty in conforming to a single authoritative standard and a single lingua franca of metadata
- Clashes with the principle of user-convenience
- Inhibition of change (e.g. the MARC standard has been difficult to change as it is deeply rooted and well-established).

2.9.5. Metadata Standards for the Semantic Web

In January 6, 1997, Tim Berners-Lee wrote a proposal entitled “Metadata Architecture” and defined metadata as “machine understandable information about web resources or other things” (Berners-Lee, 1997). It is also apparent that the use of the word metadata in the Berners-Lee's initial proposal refers to what later came to be known as the Resource Description Framework (RDF), which was subsequently approved by W3C in 2004. Berners-Lee argues in favour of maintaining the centrality of metadata. He points out that metadata about one document can occur within the document itself, can be stored separately; can include meta-metadata; can be used to describe anything; can be used either individually or in combination with other metadata (Berners-Lee, 1997). His concern was mainly of metadata to describe web resources. He was also primarily interested in defining machine processable metadata for data and documents, providing meaning and context through typed relations, which was tied to his vision of the Semantic Web as described in Berners-Lee, Hendler, & Lassila (2001).

RDF's simple data model enables the creation of semantic links among information resources (Alemu, Stevens, Ross, et al., 2012; Berners-Lee, 2009; Cobden, Black, Gibbins, Carr, & Shadbolt, 2011; Shadbolt, 2010; Styles, Ayers, & Shabir, 2008; Wallis, 2011a; Williams, 2010; Wilson 2010). RDF schema adds vocabularies, such as Class, SubClass, Domain and Range, to enable a more meaningful representation of resources. Subsequently RDFS was extended with yet additional vocabularies (Alemu, Stevens, Ross, et al., 2012). OWL allows the definition of additional semantic constructs, such as equivalency, inverse and cardinality relations and constraints (Allemnag & Hendler, 2008; W3C, 2004a). One of the defining features of the RDF model is its ability to identify resources and metadata attributes (relations) uniquely and globally using URIs. The use of URIs for metadata element names, labels and relations, according to Nilsson (2010), helps to avoid naming and identification conflicts in the use of elements.

This is also suggested by Day (Day, 2000, 2003a, 2003b) and Rothenberg (2008). Unfortunately, although there happen to be several academic papers and technical specifications regarding RDF, RDFS, SPARQL and OWL, there are, up until now, no viable Semantic Web–related metadata solutions in widespread use in libraries.

Reviewing the literature shows that current metadata standards result in metadata that is mainly attuned to human consumption rather than machine processing. This is attributed to both conceptual and technical limitations of the standards and technologies used in libraries. Coyle (2010, p.11) argues that “the library catalogue has been the sole context for library data” hence, failing to interoperate with external information providers. The challenges for principles, standards and protocols could be looked at from two perspectives: conceptual and technical. The conceptual underpinnings of contemporary metadata standards, such as FRBR and RDA, have arguably resulted in metadata records as documents attuned to human consumption rather than machine processing (Coyle & Hillmann, 2007).

Furthermore, when such metadata principles are implemented using technical formats such as MARC, the resultant records exhibit problems of metadata duplication, data inconsistency, lack of granularity and complexity (Coyle, 2010; Coyle & Hillmann, 2007; Day, 2000; Guenther & McCallum, 2003; Tennant, 2002). Even though the call for an end to the use of MARC has been proclaimed as long overdue (Tennant, 2002), the standard remains the dominant metadata structure used in libraries to this day. This can directly be attributed to several causes, including the fact that MARC is deeply embedded in library systems and functions and, thus making any changes would simply become too difficult and expensive; or that MARC, after all, is ‘adequate’ enough for libraries and serves its purposes; or it may be that alternative formats, including eXtensible Markup Language (XML), fail to deliver the additional functionality required to merit and justify the changeover. However, there are still grave doubts regarding the adequacy of MARC, espoused by several metadata experts who assert that the standard is not suitable for machine process-able and actionable metadata (Coyle, 2010; Coyle & Hillmann, 2007; Wallis, 2011a, 2011b).

Suggested alternatives to alleviate the document-centric view of contemporary library metadata approaches include the adoption of Linked Data. As its inventor, Berners-Lee, et al (2001), state, the original web has conceptually been document-centric in which the links do not carry any semantics with them and when implemented with technical formats such as Hyper-Text Mark Up Language (HTML), the resultant web pages are more attuned to human consumption, rather than machine processing. Conversely, whilst the web has been exceptionally efficient for sharing documents and creating possibilities for collaboration, a document requires human intervention for understanding its semantics once it is presented and displayed on user’s computer screen.

Put simply, machines cannot make sense of such documents. To alleviate this limitation, Berners-Lee, et al (Berners-Lee, 1997; Berners-Lee, 1998; Berners-Lee et al., 2001) came up with the concept of Linked Data (Berners-Lee, et al., 2001). The technologies to implement Linked Data include RDF, RDFS, SPARQL and OWL (Allemnag & Hendler, 2008; Berners-Lee, 1998; Decker et al., 2000; W3C, 2004a, 2004b). According to Allemnag and Hendler

(2008, p. 7) “the main idea of [Linked Data] is to support a distributed web at the level of the data rather than at the level of presentation [documents]. Instead of having one webpage point to another, one data item can point to another, using global references called URIs”.

It is worth noting that as recently as 2011, some national and regional initiatives such as the British Library, the National Library of France and the Europeana Digital Library have announced their plans to open their legacy bibliographic records as Linked Data (Williams, 2010). A report commissioned by the World Wide Web Consortium (W3C) was published in October 2011 (W3C, 2011). The report acknowledges the low uptake of Linked Data in libraries whilst providing recommendations for libraries to embrace the Linked Data principles which, among other things, includes the use of technologies such as Uniform Resource Identifier (URI), Resource Description Framework (RDF), Protocol and RDF Query Language (SPARQL) and Web Ontology Language (OWL). The report states the importance of making bibliographic library data openly and freely accessible in a form that is “shareable, extensible and easily re-usable” (W3C, 2011). Whilst the Semantic Web has not made significant strides in the library domain, its potential role for metadata encoding, representation, and sharing is indicated (Day, 2000, 2003a, 2003b; Nilsson, 2010; Rothenberg, 2008).

Despite the growing interest in these novel approaches, there is still an on-going, widespread debate as to whether incremental changes made onto traditional library-centric models and record formats, such as Resource Description and Access (RDA), Functional Requirements for Bibliographic Records (FRBR) and MACHine-Readable Cataloguing (MARC) do suffice or whether there is compelling reason for abandoning some or all of them altogether and adopting these new approaches (Coyle, 2010; Coyle & Hillmann, 2007; Dunsire 2008; Dunsire, 2009; Dunsire 2012; Dunsire & Willer, 2011; Marcum, 2011; Styles, 2009; Styles et al., 2008; W3C, 2011; Wallis, 2011a, 2011b).

2.10. Limitations of Standards-based Metadata Approaches

To discuss limitations of a standards-based metadata, one should be cognisant of whether the metadata in question is digital or paper-based (appearing on paper or card in a human-readable form). In the paper-based context, some of the limitations are inevitable whilst the move to the digital realm usually obviates most of them, e.g. placement of the same information node on different branches in the hierarchy is made possible when storage space constraints become minimal or non-existent (Weinberger, 2007). In the digital realm, Weinberger attests, information can be categorised in different and innumerable ways, based on the context and use of the information (Weinberger, 2007). To quote Weinberger (2007, p.83) at length:

“In the third order of order [in full-text digital libraries], a leaf [an information node] can hang on many branches; it can hang on different branches for different people, and it can change branches for the same person if she decides to look at the subject differently. It’s not that our knowledge of the world is taking some shape other than a tree, or becoming some impossible-to-envision four-dimensional tree. In the third

order of order, knowledge doesn't have a shape. There are just too many useful, powerful, and beautiful ways to make sense of our world.”

Whilst the role of the Online Public Access Catalogue (OPAC) has been instrumental in facilitating the identification, access and utilisation of library collections, the present state of the OPAC is criticised for failing to reflect users' search and discovery behaviours (Buchanan & McKay, 2011; Calhoun, 2005; Evans, 2009; Marcum, 2005). Evans (2009, p. 14), for instance, provides examples of OPAC systems which are unforgiving for those users that are not familiar with the terminologies used by librarians to describe the catalogue entries. According to Buchanan and McKay (2011) most users' search behaviour is characterised by the employment of unconventional metadata, approximation of author names (e.g. “I think the author name finishes with ‘Ishky’”), as well as incomplete titles (Buchanan & McKay, 2011, p.272).

However, even whilst embracing computerised, digital metadata, such as the OPAC, libraries often had a propensity to carry forward some of the limitations inherent in the old, paper-based card cataloguing system. This is mainly a consequence of libraries remaining faithful to the conventional, foundational, principles of metadata (see 2.5). For Kuhn (1962), this tendency to resist change is almost inevitable, as, according to him, institutions that religiously adhere to their conventional practises are very likely to maintain the status quo, even whilst the latter has already exhibited some manifestations of failures in addressing new and emerging problems.

One of the major inherent limitations of standards-based metadata approaches arises from the way that controlled vocabularies are constructed and utilised. These constructs, which includes thesauri, taxonomies and ontologies, are deemed, vital tools for facilitating information retrieval, as they provide pre-set terms to users. They are extensively manifested in online library databases while their use in the Library and Information Science domain has been well documented. However, the trouble with controlled vocabularies starts with the fact that their designers and/or developers cannot possibly exhaustively list all the terms available in any given domain. Furthermore, it is extremely difficult to incorporate all the subtleties and nuances pertaining to a given concept-term or the variety of its uses. This is progressively becoming more apparent in today's information landscape, wherein users of an information system can originate from different cultures and possibly have diverse interpretations and understandings of terms. Moreover, whereas, vocabularies in natural languages have a certain amount of latitude, with regard to meaning, it is stipulated that use of vocabularies in digital library systems or online information systems be strictly constrained as each is assigned a pre-defined and invariant meaning. The rationale for this convention is the assumption that use of carefully chosen and predetermined set of terms increases precision of search results. Yet another shortcoming apparent in controlled vocabularies is that, despite aiding online searches, they are also liable to become a limiting factor as the search environment is limited to entries that had been already indexed, whereas users may not know what these are.

Controlled vocabularies help in attaining consistency, by providing guidelines on the syntax and semantics of metadata values. Duval, et al. (2002) define semantics as the meaning of

metadata values while syntax is the form in which these values are recorded. According to Duval, et al. (2002), agreement on the meaning and encoding of metadata values are essential for the exchange of metadata among communities of information providers. Conforming to existing standards not only reduces cost, but also substantially fosters metadata and information resources exchange processes by enforcing consistency.

Despite all these perceived benefits, the irony is that standards themselves are in need of standards to make them talk to each other. As Caplan (2000) observes “standards development is particularly difficult within the digital library arena, primarily because the most active players have not yet formed a true community in the sense of having evolved a common vocabulary, commonality of interest, or structures for collaboration and communication.” In the choice of specific metadata standards for a particular collection, National Information Standards Organisation’s (NISO) guidelines stipulate that it is crucial to analyse the appropriateness of the prospective standard, although, in some instances, based on the nature of the collection at hand, the use of multiple standards as application profiles may also be permissible (NISO, 2007).

There exists, however, a fundamental question that needs to be addressed when choosing a specific set of metadata standards. This dilemma is whether existing metadata standards themselves adequately reflect users’ needs and expectations. This is especially important in the context of current social-technical trends, wherein users have progressively become empowered, more active players, and thus more demanding. One way of investigating the adequacy of standards, in meeting users’ needs, is re-examining the way they are devised, developed and maintained. Another viable approach would be asking the users themselves what they think of these standards. The responses to these and similar questions would help in refocusing the debate from being one that asks “Which standard to use?” to one that seeks answers to “What metadata is required for improving the findability of information objects in digital libraries?”

Costs related to the creation and maintenance of classification systems are considered major burdens of the standards-based metadata approach (Barbosa, 2008). However, perhaps the most crucial limitation of standards-based metadata approaches is the fact that “the vocabulary of the information professional/librarian may be at odds with the language of content creators and users, obscuring the very information the taxonomy should reveal” (Barbosa, 2008). Finally, the need for classifying and categorising objects within the digital world has been put in to question, as it is deemed a forced abstraction, carried over from our habits of organising physical objects in the physical world (Shirky, 2005). Thus classification systems, such as taxonomies and ontologies, as well as categorisation systems, such as Dewey Decimal Classification (DDC) and Library of Congress Classification Scheme, are seen by some as being severely limiting for organising and/or accessing information objects in the digital world (Shirky, 2005).

Going a step further, Shirky (2005) questions the assumptions that underlie the derivation of explicit, formal specifications for concepts (entities). Failure to take into account the diversity of users’ perspectives and the dynamicity inherent in their use of vocabularies, Shirky argues,

is likely to affect the findability of information objects. As Shirky (2005) further notes, “users have a terribly hard time guessing how something they want will have been categorised in advance, unless they have been educated about those categories in advance as well, while the larger the user base, the more work that user education entails.”

Buckland (1999) pointed out, the challenges of the mismatch between the vocabularies of authors, librarians and indexers as opposed to searchers. Thus, emphasising the centrality of incorporating the various vocabularies and interpretations involved in the description of information objects. Buckland ascribes the challenges partly to linguistic and cultural differences. In libraries, the development and use of controlled vocabularies are mostly grounded on deterministic assumptions of users’ vocabularies (Shirky, 2005; Weinberger, 2007). Chiefly stipulated top-down, controlled vocabularies embody fixed categorical elements that are updated at lengthy intervals. Consequently, the terminologies are likely to be outdated, missing context over time, and hence prone to failures in adequately representing users’ world views. Furthermore, as Veltman (2001) notes, while librarians are preoccupied “with a controlled grammar,” users may not have “any idea at all of [this] grammar.”

Libraries have, hitherto, focused on standards-based metadata architectures with pre-determined schema (constraints and guidelines) and their associated controlled vocabularies and authority lists (Alemu, Stevens, et al., 2012b). In their book “The Social Life of Information”, Brown and Duguid (2000), contend that institutions in this information age tend to adopt a top-down view of technological determinism, fuelled by the likes of Moore’s Law, whilst ignoring the socio-technical aspects of information technology. Brown and Duguid (2000, p. 1) argue that ignoring the social facets of information leaves out those aspects of information technology use which “lies around the edges, context, background, history, common knowledge and social resources”. Libraries, as institutions that utilise information technologies, are not immune from being confronted with these challenges as they endeavour to effectively incorporate the about-ness of information objects (Buckland, 1999). As Buckland (1999) notes, an “unfamiliar vocabulary reduces search effectiveness”, indicating thus the importance of incorporating the terminologies from the users themselves, who are the *raison d’être* for libraries.

2.11. Conclusions

As this review of the related literature shows, current metadata approaches are based upon agreed upon principles including the principle of sufficiency and necessity, the principle of user convenience, the principle of representation and the principle of standardisation. Most of these principles result in metadata schemas with fewer metadata elements, which in turn results in metadata simplicity. Most of these principles are built on assumptions that standardisation provides efficiency in metadata creation and management. The focus on standardisation is also hinged upon the assumption that metadata conformance and uniformity (through controlled vocabularies and encoding schemes) would bring better interoperability and metadata sharing. Whilst, standardisation and expert created (controlled) metadata has its benefits, as the literature review shows, it fails to adequately represent the diversity of views and perspectives that exist in library users. Hence, standards-based

metadata approaches have come under criticism for being rigidly hierarchical and authoritative (Alemu, Stevens, & Ross, 2012b; Shirky, 2005; Weinberger, 2007). Current metadata principles exhibit limitations in light of changing users' needs, the existence of multiple interpretations and changes in technological trends, such as social media and Web 2.0.

It is thus important to look into how these issues can be addressed. The following chapter reviews related literature on the potential role of the Web 2.0 paradigm for metadata creation and utilisation.

Chapter Three: The Web 2.0 Paradigm and the Emergence of Socially-Constructed Metadata Approaches

3.1. Chapter Overview

In order to provide background and context, this chapter reviews related literature on the role of the Web 2.0 paradigm for libraries and the emergence of socially-constructed metadata approaches. Even though, the Web 2.0 paradigm is often discussed in the metadata literature at the level of technologies (such as user tagging, ratings, reviews and recommendations), the interest in this thesis is on the conceptual underpinnings rather than the specific technological implementations (instantiations). The literature thus focuses on Web 2.0 concepts including two-way collaboration, users as co-creators, the “Wisdom of Crowds”, variable participation, openness and post-hoc quality control. However, to support the discussion with examples, some applications are also mentioned in this review. Thus, the first section of the review focuses on Web 2.0 concepts and the latter on Web 2.0 technologies and applications. Finally, the chapter indicates the implications of the Web 2.0 paradigm for libraries, in particular to library metadata.

3.2. Web 2.0 Concepts

As the dust from the dot-com hype and bubble began to settle after the year 2000, Dale Dougherty and Tim O'Reilly examined those common factors that uniquely enabled some web-based businesses to sustain themselves and prosper, whilst many others had gone out of business (O'Reilly, 2005). O'Reilly (2005) noted that “far from having ‘crashed’, the Web was more important than ever (O'Reilly, 2005). The success of the Web 2.0 paradigm, Floridi notes, is its focus on people (participation) and metadata. He points to the importance of differentiating between semantic intelligence (truth and understanding) and metadata (identifiable information). For him, “humans are the only semantic engines available”. It is also indicated that the Web 2.0 paradigm is predominately a socio-technical phenomenon (Lagoze, 2010; Miller, 2005; O'Reilly, 2005), thus as O'Reilly (2005) and Miller (2005) contend, it has more to do with attitude and culture than technology. The participatory nature of Web 2.0 (O'Reilly, 2005) and its focus on metadata (Floridi, 2009) has implications for education in general (Alexander, 2006; Anderson, 2007; Nielsen, 2012) and libraries in particular (Alexander, 2006; Casey & Savastinuk, 2006; Evans, 2009; Kroski, 2008; Miller, 2005).

O'Reilly identified the following concepts underpinning the Web 2.0 paradigm. These include active two-way collaboration, users as co-creators, the Wisdom of Crowds, variable participation and openness (O'Reilly, 2005). These concepts were further elaborated and discussed by several other authors (Alexander, 2006; Anderson, 2007; Casey & Savastinuk, 2006; Evans, 2009; Floridi, 2009; Kroski, 2008; Maness, 2006; Miller, 2005; O'Reilly, 2005; O'Reilly & Battelle, 2009).

3.3. Two-Way Collaboration

Web 2.0 technologies allow and encourage collaborative participation of users/customers surpassing time and spatial boundaries. One of the central tenets of the Web 2.0 paradigm is

its capability to facilitate collaboration, coordination and two-way communication (Anderson, 2007; Floridi, 2009; O'Reilly, 2005). In his critique of the Semantic Web versus Web 2.0, Floridi notes the problematic issues of ascribing intelligence and semantics to machines, asserting that “for the foreseeable future, the responsibility for such a gigantic task [of intelligence and semantics/understanding/meaning] will remain totally human” (Floridi, 2009). Whilst the Semantic Web is primarily intended to provide machine process-able data, Web 2.0 fosters the tools for people to collaborate (Floridi, 2009).

In Floridi's view, unlike the Semantic Web, Web 2.0 lacks clear definition and specification, albeit its potential to bring real intelligent beings, i.e. humans, together to collaborate and solve problems in coordination (Floridi, 2009). Similarly, in his book, “Re-inventing Discovery: the new era of networked science”, Nielsen (2012) argues that open and collaborative approaches such as blogging, if implemented and managed properly for example with appropriate motivation and incentives to scientists, would amplify the “collective intelligence” of humans. Nielsen (2012) maintains that Web 2.0 collaboration has the potential to serve as the “architecture of attention” (identification of special expertise). He thus argues that the use of Web 2.0 and social media for serious scientific projects is not a “frivolous” waste of time.

3.4. Users as Co-Creators

One of the central concepts of the Web 2.0 paradigm is the notion of involving users as co-creators of content and metadata (O'Reilly, 2005). This conceptual underpinning assumes two-way collaborations between users and institutions, such as libraries, and has a strategic advantage. This emerging strategy has helped to realise what Toffler (1980), in his book “The Third Wave”, envisioned as the shift from passive consumerism to proactive prosumerism. In it, Toffler saw that two-way collaborative approaches, fostered by the social and political will, reoriented the traditional ways of doing business and offering services. As Kroski (2008) notes the Web 2.0 paradigm considers users as major stakeholders whose proactive participation is sought as a central strategic competitive advantage. Kroski adds that Web 2.0 is inclusive by design where users can “participate, organise, read, write and play online”. It is thus important to explore the implication of the concept of “users as prosumers” in library metadata functions. One of the emerging concepts in light of users as prosumers is the notion of the “Wisdom of Crowds”, where emphasis is placed on the collective and aggregate value that can be harnessed from users rather than the individual contributions.

3.5. The Wisdom of Crowds

One of the concepts of the Web 2.0 paradigm is the “Wisdom of Crowds” (Surowiecki, 2004). Employing the Web as a platform, O'Reilly argues, the small contributions made by individual users can be collectively re-mixed, aggregated and harnessed, providing thus what he refers to as collective intelligence (O'Reilly, 2005). Principles that help in harnessing collective intelligence include long-tail distribution of products and services, including talent across the network; low marginal costs of cooperation and coordination; granular addressability of content; and the network effect, the architecture of participation and the

concept of users as co-creators are central pillars of collective intelligence (Anderson, 2006; O'Reilly, 2005; Shirky, 2008).

The phrase “Wisdom of Crowds” was popularised by James Surowiecki, in his oft-cited book “The Wisdom of Crowds: Why the Many are Smarter than the Few” (2004). Surowiecki describes how small, but unique, contributions aggregate and become greater than the sum of the constituent parts. Surowiecki (2004, p.28) emphasises the importance of diversity, independence of thinking, and activity-as-a group for adding value in the form of collective wisdom. Surowiecki also notes the importance of decentralisation and specialisation (diversified knowledge base) in enhancing the value of contributions. Surowiecki (2004, p.30) argues that “cognitive diversity needs to be actively selected, and it’s important to do so because, in small groups, it is possible for a few, biased individuals to exert undue influence and skew the group’s collective decision”. The goal is not to strive for consensus, homogeneity, and group-thinking in knowledge creation or decision making, Surowiecki argues. Most importantly, the theory of the “Wisdom of Crowds” posits that the idea of gathering diverse groups of people does not preclude the inclusion of the smartest person. Surowiecki admits that, just as democracy has its own limitations, emanating from the fallibility of its enactments, so does the “Wisdom of Crowds”. Harnessing the collective intelligence of groups of people involves well-thought out coordination, cooperation as well as aggregation costs.

3.6. Variable Participation

An integral component of the concept of the “Wisdom of Crowds” is the phenomenon known as ‘variable participation’ (Shirky, 2008), whereby no contribution is considered too small, but is aggregated to form collective intelligence (O'Reilly, 2005). In other words, it is not the case that every contributor adds or edits a complete entry, but that the numerous ‘tiny’ contributions slowly accrue to bring about a network effect, the phenomenon wherein value appreciates as the number of users grows. Most proponents of the Web 2.0 paradigm are of the same opinion, stating that, unlike the conventional peer-review model, where individual contributions make a significant mark, the Web 2.0 paradigm functions in a bee-hive or ant-colony fashion, where the contribution of a single bee is insignificant without taking the sum of contributions of the whole hive (Shirky, 2008). Wright (2007, p.14) notes that “no individual bee possesses the intelligence to make such a decision, but as a group, the bees generate a collective ‘mind’ far cleverer than the sum of its tiny-brained parts.”

3.7. Openness

Reviewing the metadata literature shows that, one of the core concepts of the Web 2.0 paradigm is that its architecture facilitates participation (O'Reilly, 2005), which has lowered the barrier to entry for contributors. Central to this architecture of participation is the importance of embracing openness to sharing and collaboration, based on mutual trust (Alexander, 2006; Anderson, 2006, 2010; Shirky, 2005; Tapscott & Williams, 2010; Udell, 2004; Weinberger, 2005, 2007). In other words, the Web 2.0 paradigm is characterised by the participatory and collaborative culture which is built in around its architecture, enabling users to become proactive content creators and consumers. In their bestselling book “Wikinomics”,

Tapscott and Williams (2010) assert that “due to the deep changes in technology, demographics, business, and the world, we are entering a new age where people participate in the economy like never before”. They identify openness as one of the major principles underlying mass collaboration, along with peering (self-organised peer networks of contributors), sharing and acting globally, each of which extends the possibilities for tapping into a much larger pool of talent. They also note that “openness is associated with candour, transparency, freedom, flexibility, expansiveness, engagement and access”. According to them, current economic, social and technological trends suggest that openness does not necessarily correlate to intellectual property infringements. They further contend that the culture of openness, and continued recognition of its potential benefits, has compromised the “conventional wisdom that says companies compete by holding their most coveted resources close to the chest” (Tapscott & Williams, 2010).

Alexander (2006) asserts that the multi-directional flow of information, between producers and consumers and across domains, servers and machines, necessitates opening up information silos and fostering shared services. The author contends that even commercial sites, such as Amazon.com, permit their users to “harvest ISBN numbers from its listings”. He attests that “openness remains a hallmark of this emergent movement, both ideologically and technologically.” Contextualising this to libraries, Miller (2005) argues that the principles of Web 2.0 are predicated on the notion of liberating data, which in turn allows data to be “exposed, discovered, and manipulated” in a multitude of ways, thereby creating unimaginable possibilities for re-purposing and re-using the data. In terms of recent technologies, whilst Linked Data can be made usable without it necessarily being open (Cobden, Black, Gibbins, Carr, & Shadbolt, 2011; Shadbolt, 2010; W3C, 2011), as Berners-Lee (2010) emphasises, opening data brings forth numerous benefits to society. In his TedTalk, Berners-Lee (2010) re-iterates his vision of Linked Data by citing a number of international, regional and community-based initiatives and projects that have adopted Linked Data principles, and thereby made their data openly available using URIs and RDF technologies. As Berners-Lee (2010) re-emphasises, open data can be re-used in an unimaginable number of ways. For instance, open government data enables tax payers to check how, where and for what purposes their money has been spent by their elected representatives. As Berners-Lee (2010) stresses the momentum for opening up data “has only just started”. According to Kroski (2008) the concept of openness facilitates re-usability, mash-ability and re-mixability of information and metadata created at disparate locations.

3.8. Post-Hoc Quality Control

Web 2.0 proponents, such as Shirky (Shirky, 2005, 2008, 2010), and Weinberger (2007), argue that Web 2.0 content is neither anarchic nor “anything goes”. However, unlike the contemporary (print or Web 1.0) model where the editing, vetting and the quality control process is performed before content is published; the Web 2.0 paradigm implements the process of filtering “after-the-fact (post-hoc)”. According to Shirky (2005), “There’s an analogy here with every journalist who has ever looked at the Web and exclaimed ‘Well, it needs an editor.’ The Web has an editor, it’s everybody. In a world where publishing is expensive, the act of publishing is also a statement of quality; the filter comes before the

publication. In a world where publishing is cheap, putting something out there says nothing about its quality. It's what happens after it gets published that matters. If people don't point to it, other people won't read it. But the idea that the filtering is after the publishing is incredibly foreign to journalists.” Shirky (2005) furthermore contrasts the new model with library cataloguing, where classification schemes and categories are pre-determined.

3.9. Web 2.0 Technologies and Applications in Libraries

Whilst Web 2.0 technologies (such as tagging, social bookmarking, reviews, blogging, crowd sourcing and recommendations) are instantiations of the concepts that underpin the Web 2.0 paradigm, applications are the specific tools developed to implement the instantiations. The Web 2.0 concepts are hinged on positioning the Web as a collaborative, participatory, and strategic technological platform (Anderson, 2009a, 2009b; Anderson, 2007; Floridi, 2009; Maness, 2006; Miller, 2005; Morville, 2005; O'Reilly, 2005; Weinberger, 2005, 2007). According to O'Reilly, companies that capitalised on these Web 2.0 concepts have weathered the dot-com bust and remained competitive having taken advantage of emerging trends in Web 2.0 (Anderson, 2006; O'Reilly, 2005; Shirky, 2008; Tapscott & Williams, 2010).

According to O'Reilly (2005), Web 2.0 technologies constitute a major technological shift from the stand-alone and installable applications of Web 1.0 technologies that are made available as services (O'Reilly, 2005). Web 2.0 applications such as Wikipedia, Flickr, Delicious and LibraryThing (an online collaborative social cataloguing website) (Casey & Savastinuk, 2006; Farkas, 2007; Maness, 2006; Miller, 2005) facilitate bottom-up collaborative efforts to create content and metadata.

With the adoption of Web 2.0 technologies, photo and video sharing sites such as Flickr and YouTube come to the scene. Nielsen (2012) cites successful Web 2.0 applications used for scientific projects such as Tim Gower's blog (to solve mathematical problems collaboratively), Wikipedia and Galaxy Zoo (citizen science project for discovering planets and stars). Nielsen also acknowledges that the adoption of Web 2.0 technologies in academia is yet in its early development stages, and suggests that institutions and scientists should fully embrace the Web 2.0 paradigm and online scientific collaborations for the good of science and society. Nielsen further contends that current institutional incentives should be re-considered so as to promote knowledge sharing and collaboration. It is important to note that Nielsen's arguments are equally valid to the library environment.

Web 2.0 technologies such as tagging and reviews have also been implemented by some libraries, albeit as add-on (Farkas, 2007; Smith, 2008). The new technologies allow users to tag the resources they access and use. Most of the users do tagging for their own purposes but there are also people who tag out of the urge to help others to find the resources. The term used for social tagging is folksonomy. Web 2.0 applications such as Delicious, LibraryThing, Diigo and Flickr allow their users to bookmark information objects such as photographs, books and articles (Smith, 2008).

The actual and potential benefits of Web 2.0 technologies for libraries are discussed at length by several authorities (Casey & Savastinuk, 2006; Evans, 2009; Maness, 2006; Miller, 2005;

Smith, 2008). The adoption of the Web 2.0 paradigm in libraries is mainly at the level of the application wherein such applications developed by commercial or third party companies (e.g. Flickr, LibraryThing, Facebook and Twitter) are brought in as add-ons, whilst standards-based (traditional) library functions continue to operate without any significant changes.

According to Smith (2008) social tagging provides users with an easy-to-use, flexible, extensible and aggregatable metadata ecology, which in no small measure augments the return on experience for users, as well as the return on investment for institutions. Such return comes in various forms, including enhancing findability, increasing user participation, facilitating collaboration, relatively cheap means of obtaining descriptive data, identifying patterns of information of use behaviour, such as popularity of sources, users' interests and emerging vocabularies (Smith, 2008).

According to Gruber (2007, p. 3):

“Google's revolution in search quality began when it incorporated a measure of ‘popular’ acclaim - the hyperlink - as evidence that a page ought to be associated with a query. When early webmasters were manually creating directories of interesting sites relevant to their interests, they were implicitly ‘voting with their links.’ Today, as adopters of tagging systems enthusiastically label their bookmarks and photos, they are implicitly voting with their tags. This is, indeed, ‘radical’ in the political sense, and clearly a source of power to exploit”.

Tagging technologies allow different users to assign dissimilar tags to the same information object. For instance, as Alexander (2006, p. 34) notes, “a historian photographs the Waterloo battlefield, uploads the result to Flickr or 23, and adds keywords meaningful to her: Napoleon, Wellington, Blucher, 1815. A literature scholar creates similar images but tags them according to his interests: Thackeray, Hugo and Clarke”. For many social media proponents this is considered a new way of creating metadata (Anderson, 2006; Shirky, 2005; Weinberger, 2005, 2007).

As Weinberger (2005) explains tagging is a major shift in cataloguing in that not librarians, not authors but for the first time readers are able to describe objects. Weinberger argues that “an author is an authority when it comes to what she intended her work to be about, but not about what it means to others. When it comes to searching, what a work means to the searcher is far more important than the author's intentions.” In his book “Everything is Miscellaneous”, Weinberger (2007, p.32) cites several examples about the arbitrariness of categorising objects and argues that there is no one correct way of ordering as each scheme is deeply intertwined with the social, political and cultural realities of the categoriser. Similarly, Shirky (2005) argues that the claim to have one ontological structure that is true for everyone is a “platonic ideal”. Both argue that, when it comes to findability, it is the readers' intention that should count. Weinberger (2007) contends that the author provides metadata from his/her own perspectives and the multitude of metadata that enriches the information object should come from users. Hence, both Shirky and Weinberger argue that social tagging answers part of the problem of traditional taxonomies and classification systems (2.10).

Both Weinberger (2005) and Shirky (2005) also maintain that the attempt to describe objects under pre-set categories as well as expecting users to search for them employing those pre-set keywords, is flawed. Weinberger states, “Tagging repudiates one of the deepest projects our culture has undertaken over and over again: the rendering of all knowledge into a single, universal framework. This rendering has been assumed to be an aid to the process of discovery: the universe has an inner order that experts and authorities can expose. But in a networked world we know better than ever that such an order is a myth of rationality”. As Weinberger continues, humans are simply unable to agree on a single structure. This, Weinberger says, has become even more apparent with the chaotic nature of tags. As Shirky (2005) also argues, though the absence of structure in tagging systems at first may seem a “recipe for disaster, but as the Web has shown us, you can extract a surprising amount of value from big messy data sets.”

Acknowledging the importance of social-metadata, Lu, Park and Hu (2010) compared a selected set of user tags found on a social cataloguing website (LibraryThing) against expert-assigned subject terms using the Library of Congress Subject Headings (LCSH). A total of 8562 book records were contrasted. Findings revealed that, in the social metadata system, users had applied 176,105 unique tags while a total of only 7,628 terms had been applied by librarians using the LCSH, indicating a large disparity in number of terms between librarian-created versus user-created metadata. According to Lu, et al. (2010), the large number of users participating in social cataloguing systems is the main factor for this very significant difference in size. It is important to note that librarians provide only one set of metadata, in contrast to the diversity of metadata that comes from users participating in such Web 2.0 environments. As Lu, et al. (2010) found, librarians assigned a maximum of five subject headings (terms) to an individual book, in contrast with a maximum of 3909 tags assigned by users for a single book.

One of the most interesting findings in the study by Lu, et al. (2010) is that the overlap between tags used for describing books in LibraryThing and terms in LCSH was just 2.2%, indicating another disparity between the two. Even though accuracy and formality are crucial elements of paramount importance for standards-based metadata approaches, there usually are emergent vocabularies, used in regular day-to-day exchange among searchers, which might not have been incorporated into these controlled vocabularies. For instance, a user may use the keyword ‘Mac’ to search for the Macintosh personal computer. Hence, tagging systems allow users to describe information objects in their own words.

However, Lu et al. (2010) argue controlled vocabularies are important for ensuring metadata consistency. In addition, they point out that controlled vocabularies, such as the LCSH, allow granular description of information objects, which can be attributed to the hierarchical structure inherent in them. Lu et al. (2010) also note the limitations of tags, for instance some tags are very personal (“to read”) and/or idiosyncratic. In addition, they also noticed that social metadata systems suffer from an absence of synonym as well as homonym control (Gruber, 2007; Smith, 2008).

Mathes (2004) and Merholz (2004) contend that socially-constructed metadata approaches can be used in the design of formalised metadata systems, such as controlled vocabularies. Mathes (2004) refers to user-generated metadata as cooperative-classification; whilst Merholz (2004) dubs it ethno-classification. Both authors agree that user-generated metadata scales better than traditional standards-based hierarchies and is better at providing pathways to serendipitous discovery of information objects. Mathes (2004) points out the importance of analysis and aggregation of user-generated metadata for generating more meaningful metadata that reflects the large diversity of views. In this regard, librarians and metadata system designers of OPACs (the discovery interfaces) would do well to learn from what are known as desired lines (Merholz, 2004). For maximising the benefits from user-contributed metadata, Merholz (2004) argues that a lot can be learned from desired lines, which are analogous to “trails [that] demonstrate how a landscape’s users choose to move, which is often not on the paved paths. A smart landscape designer will let wanderers create paths through use, and then pave the emerging walkways, ensuring optimal utility”. ‘Desired lines’, being emergent ethno-classification alternatives, can help guide the creation and maintenance of controlled vocabularies. However, ‘desired lines’ are not used in current metadata systems, possibly because contemporary standards-based approaches are top-down and hierarchical.

In summary, the Web 2.0 paradigm presents a new opportunity for librarians and users to converse and collaborate (Casey & Savastinuk, 2006; Evans, 2009; Kroski, 2008; Maness, 2006; Miller, 2005). With the growth in popularity of the Web 2.0 paradigm, libraries, have come to realise, albeit slowly, its potential benefits (Evans, 2009; Farkas, 2007; Guy, 2006; Kroski, 2008; Maness, 2006; Mathes, 2004; Smith, 2008; Udell, 2004; Weinberger, 2005, 2007).

As the metadata literature indicates there is recognition of the potential role of Web 2.0 technologies and applications for libraries (Evans, 2009; Farkas, 2007; Guy, 2006; Kroski, 2008; Maness, 2006; Mathes, 2004; Smith, 2008; Udell, 2004; Weinberger, 2005, 2007). Some consider tagging as a radical new approach to creating metadata (Shirky, 2005; Weinberger, 2005). It is also important to note that current literature on metadata seems to focus on tagging, which is considered the predominant form of user-generated metadata. In addition more emphasis is placed on implementing one or more of the Web 2.0 applications such as LibraryThing, Flickr, StumbleUpon, CiteULike, Connotea, Diigo, GoodReads, Shelfari, Delicious and SecondLife (Kroski, 2008). In general, the implementation of Web 2.0 concepts, technologies and applications of the Web 2.0 paradigm has several implications for libraries. The following case of Wikipedia versus Britannica presents an interesting example where it shows the implication of the effective implementation of Web 2.0 concepts rather than the mere applications, thus indicating the importance of considering the conceptual underpinnings.

3.10. The Case of Wikipedia versus Encyclopaedia Britannica

The case of Wikipedia is one of the most-oft cited examples of content created through the “Wisdom of Crowds” (Anderson, 2007; Shirky, 2007; Surowiecki, 2004; Weinberger, 2007).

In his book, *the Long Tail*, Anderson (2007) identifies three broad models for creating encyclopaedias.

The first model is the ‘single expert model,’ wherein one authoritative reference for general knowledge is compiled by one person, with For example, Aristotle’s “Encyclopaedia of Greek knowledge”, Pliny the Elder’s “*Naturalis Historia*,” and Diderot’s “*Encyclopédie*,” being famous examples. This model, as Anderson (2007) argues, is no longer feasible, as the body of human knowledge has expanded so vastly that it is almost unimaginable that a single person could possibly master all of it, even if he/she were as multi-talented as Michelangelo, the Italian genius.

The second model is a ‘team-of-experts model,’ where a selected team of scholars assemble the reference work, as is the case for *Encyclopaedia Britannica* (Anderson, 2006). These two models supposedly enforce stringent criteria in the processes of identifying, selecting, editing and reviewing entries prior to the actual publication, which may be referred to as before-the-fact editing and reviewing (Anderson, 2006).

The third model is referred to as the ‘open collective model,’ by Anderson (2006). This is a bottom-up, collaborative model, and is exemplified by Wikipedia (Anderson, 2006). Content in Wikipedia is edited and reviewed after-the-fact, i.e. after the subject matter has been published. In the third model, the ‘barriers-to-entry’ when creating an entry in the encyclopaedia are very low, while the encyclopaedia itself enforces very rudimentary structures only. Christakis and Fowler (2011, p. 279) have identified three categories of people who either create, edit or use Wikipedia: ‘Co-operators’ are people who, with good motives, create entries; ‘Free riders,’ who, although constituting the majority, do not add new entries or edit existing ones; and ‘Punishers’, who are volunteers who do the monitoring and patrolling work, such as editing entries and remove malicious ones. Fostered by diminishing costs in digital publishing and storage, the ‘open collective model’ does not put any limits as to the breadth and depth of entries. The first two models fail to scale while the third is able to do so, being capable of incorporating as many entries as there are contributors.

To test the validity of this perception, a widely cited article, “Internet encyclopaedias go head to head” (Giles, 2005), compared a total of 50 entries, in the realm of science, that appeared in both the *Encyclopaedia Britannica* and Wikipedia. The pairs of entries were sent for blind review by experts, wherein the latter were asked to identify factual errors, critical omissions and misleading statements in each pair. Forty-two of the entries were examined by the reviewers and the results revealed that “the difference in accuracy [between the two] was not particularly great: the average science entry in Wikipedia contained around four inaccuracies; in [Encyclopaedia] Britannica, about three,” indicating that malicious entries are “not the rule but exceptions”. As pointed out in the study, the novel features offered by social metadata, whereby anyone, anywhere can add and edit entries, has given Wikipedia a competitive edge over *Encyclopaedia Britannica*. Moreover, the study goes further arguing that Wikipedia benefits from having a wide spectrum of contributors, an increasing volume of entries, more frequent updates and an in-built mechanism for resolving disputes. These conclusions, though, did not go unchallenged (*Britannica Inc.*, 2006). In a rather spirited rebuttal of the

study, Britannica Inc. (2006, p. 7) refuted the findings, stating that: “the study was so poorly carried out and its findings so error-laden that it was completely without merit.”

The Wikipedia versus Encyclopaedia Britannica contention goes beyond the question of print versus online version or free versus fee-based access. As mentioned earlier, the two follow completely different models. Britannica follows a rigorous solicitation of articles from notable authorities in a specific field of interest and each contribution passes through a rigorous editing and reviewing process. Britannica also has limits on the subject matter and coverage of articles. It offers a read-only (Web 1.0) service, where once published, it is difficult to make changes as frequently as desired. Most importantly, Britannica follows an authoritative and closed approach, where no reader is allowed to write or edit articles at his or her own whim.

On the contrary, Wikipedia follows an open model and covers unlimited subject areas. Any one, from anywhere can create new entries as articles, edit them and start conversations about them with other contributors. Nonetheless, Wikipedia itself is not totally anarchic, as it has in place a mechanism, albeit minimal, for monitoring and revising entries as well as for removing articles that are deemed malicious, wrong or unacceptable to the community of users. According to Tapscott and Williams (2010) “peer production is emerging as an alternative model of production that can harness human skill, ingenuity and intelligence more efficiently and effectively than traditional forms”. This new model of production operates in a different set of motivations, including altruism, experience, community development, fun and professional responsibility, thus indicating the importance of considering non-monetary motivations. The structure and hierarchy is very minimal, while the network of volunteers is self-organising (Tapscott & Williams, 2010).

One of the core strengths of the Wikipedia model is collaboration. According to Shirky (2008), the collaboration model of Web 2.0 is built-in into the basic technological architecture itself, i.e. the platform itself facilitates collaboration, - aggregating and making sense of individual contributions. In his oft-cited book “Here Comes Everybody: The power of organising without organisations,” Shirky (2008, p.21) points out that “by making it easier for groups to assemble and for individuals to contribute to group effort without requiring formal management, [Web 2.0] tools have radically altered the old limits on the size, sophistication, and scope of unsupervised effort.” Shirky is not so naive as to assume that management is not entirely unimportant; however, he argues that the costs of organising are distributed within the network itself, and, hence this has opened new avenues for group collaboration as well as for the aggregation of these efforts.

The Wikipedia model is neither devoid of problems nor without its share of critics. It is often asserted that encyclopaedias created by the first two (expert-led) models are more credible, as entries are pre-filtered, passed through rigorous edits and are peer-reviewed (Keen, 2007). Keen strongly criticises the democratisation of media and the amateurism of expertise. According to Keen, it is gambling at humanity’s real peril to stand aside and watch Web 2.0 demolishing what has for so long been a credible and authoritative media, one that informs

and educates. However, it can be counter-argued that even the supposedly peer-reviewed and pre-filtered sources, such as Encyclopaedia Britannica, are not always accurate.

3.11. Limitations of the Web 2.0 Paradigm

The Web 2.0 paradigm is, in general, criticised for a lack of editorial control, which used to be a main characteristic of the conventional, peer-reviewed, edited, and filtered-before-publication model. In his book, “The Cult of the Amateur: How Today’s Internet is Killing Our Culture and Assaulting Our Economy”, Andrew Keen provides a book-length critique of the Web 2.0 paradigm. In the book, Keen identifies the challenges associated with Web 2.0, especially the absence of clear delineation between reader and author or fact versus fiction, which he argues is making it difficult to determine reliability, credibility and authority (Keen, 2007, p.27). Without rigorous fact-checking, editorial control and filtering-before-the-fact and ring fencing, Keen warns against the disappearance of truth and facts, on the basis of which informed decisions can be made. As Keen (2007, p.30) notes “Whereas there may be infinite typewriters, there is a scarcity of talent, expertise, experience, and mastery in any given field. Finding and nurturing true talent in a sea of amateurs may be the real challenge in today’s Web 2.0”. For Keen the uncontrolled Web 2.0 paradigm is making the intermediary professions of reviewing and editing less important. Critiquing the Long Tail, Keen (2007) points out difficulties in differentiating the good from the bad, within the unguarded user-created content. Keen (2007) contends that the democratisation of content creation, fostered by the Web 2.0 paradigm, has significant downsides. One of the chief problems, he notes, is the difficulty of identifying contributors, citing cases of malicious entries in Wikipedia, created by unidentified users (Keen, 2007, p.20). Keen (2007, p.19), extending his criticism, states that “in a flattened, editor-free world, where independent video-graphers, podcasters, and bloggers can post their amateurish creations at will, and no one is being paid to check their credentials or evaluate their material, media is vulnerable to untrustworthy content of every stripe”. This contention is not completely rejected by proponents of the Web 2.0 paradigm. They admit to the fact that, when systems like this are open to millions of people for addition and editing, it is highly likely that some entries would be inaccurate. Anderson (2006), however, counter-argues that the problems regarding user-created content can be rectified by the community of users themselves, thus creating a self-regulating, self-correcting and self-healing system.

In accordance with the above, the most-often cited challenges of the potential roles of the Web 2.0 paradigm for metadata creation and utilisation include concerns of ambiguity, the idiosyncratic nature of user-generated metadata (characterised by personal or eccentric tags), absence of synonym/homonym controls, and hence metadata quality issues (Guy, 2006; Mathes, 2004; Smith, 2008; West, 2007).

Though the Web 2.0 applications are popular, they are not without their limitations. Such limitations include: the lack of quality control of the tags, idiosyncratic and personal tags (such as ‘to be read’, ‘to buy’ or ‘gifts from mum’), and lack of structural hierarchy (broader/narrower/related terms) (Guy, 2006). For example, as used by the Flickr application,

the term ‘Apple’ can refer to any edible fruit, the Forbidden Fruit in the Bible, a computer brand, or an abbreviated form of the place known as Apple Valley, or Apple Records.

3.12. Introducing Socially-Constructed Metadata Approaches

Current adoption of the Web 2.0 paradigm focuses at the application level whereby specific third party applications such as Flickr, LibraryThing, Delicious and GoodReads are adopted as add-on without the full consideration of the concepts that underpin these applications (Farkas, 2007). Currently, the emphasis is on tagging (Smith, 2008; Weinberger, 2005, 2007), which can be seen in the adoption of the term folksonomy. The term folksonomy is credited to Thomas Vander Wal (2007) who defines it as “the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval”. The term is widely used in the metadata literature (Barbosa, 2008; Gruber, 2007; Guy, 2006; Kroski, 2005; Mathes, 2004; Peterson, 2006; Pink, 2005; Pressley, 2005; Sturtz, 2004; Vander Wal, 2007; West, 2007).

However, in this thesis, instead of the term folksonomy, the phrase “socially-constructed metadata” is introduced, as it better reflects not only tagging but also the incorporation of other facets of user-created metadata such as user reviews, ratings and recommendations. Thus, concepts in relation to socially-constructed metadata include the notion of collective intelligence (O'Reilly, 2005), the Wisdom of Crowds (Surowiecki, 2004) and the Long Tail (Anderson, 2006).

In this context, it is important to note that in the early 1930s, Paul Otlet espoused the importance of incorporating the social space of documents as part of the library cataloguing system. After reviewing the efforts of Melville Dewey's DDC and Panizzi's cataloguing principles, Otlet came to conclude that such tools only guide the reader to the location of the book but not to its contents or to the relationships between documents (Wright, 2007). Otlet thus envisioned a system called the “réseau,” a tool to create semantic links between documents and keep track of the annotations made by readers, eventually forming new trails of documents, which he calls “the book about the book” (see Otlet's original in French *Traité de Documentation*, 1934, reviewed by Wright, 2007). Paul Otlet's vision of the social space of documents has not been fully realised up until now.

3.13. Conclusions

The review of literature shows that at present standards-based metadata and socially-constructed metadata approaches are mostly considered separately, with each having its own strengths and weaknesses. In summary, standards-based approaches (Chapter Two) result in a relatively higher quality metadata, have defined categories and present well-structured schema (Zeng and Qin, 2008). However, standards are overly hierarchical, complex and fail to scale as the size of collections grows (Shirky, 2005; Veltman, 2001; Weinberger, 2007). On the other hand, the Web 2.0 paradigm presents a potential opportunity to perhaps obviate some of the challenges of current standards-based metadata approaches. Shirky (2005), for example, suggests the possibility of accommodating users' terminologies.

However, user-generated metadata, in general, doesn't have in place any systematic process for ensuring the quality and reliability of metadata; nor does it present any vocabulary (synonym, homonym) control, and, hence, appears to be chaotic and unwieldy. Resistance against the practice of user tagging, user-reviews and ratings may be attributed to the criticisms directed against the Web 2.0 paradigm in general, which include a lack of structure and an absence of editorial quality, authority and credibility (see Gorman, 2005; Keen, 2007). However, it can be argued that socially-constructed and standards-based metadata approaches are not in opposition.

The Web 2.0 paradigm encourages the co-creation of metadata by users, and thus provides what could be considered socially-constructed metadata. Some of the limitations of standards-based metadata approaches identified in this research such as the disparity in language between the librarian and the user might partly be addressed. Multitudes of interpretations of information objects can be included and hence local and individualised interests are also addressed. But the benefit of the Web 2.0 paradigm begins to show when one considers the metadata that comes from various users, communities and contexts is aggregated and analysed (O'Reilly, 2005).

In the metadata literature there are three broad positions with regard to the use of Web 2.0 technologies within libraries. Firstly, there are those that advocate a fully-socially-constructed metadata approach. Authors like Shirky (2005) and Weinberger (2007) argue that current standards-based metadata systems are broken, and hence, they suggest, should be substituted by an open, democratic, and socially constructed metadata approach. The digital world, Shirky (2005) argues, is a radical break from the print-and-analogue paradigm, as a single information object can now be categorised in an infinite number of places using hyperlinks. Shirky's argument about the inherited limitation of categorisation from the physical world is also echoed by Weinberger (2007). Supporting the use of Web 2.0 technologies, Weinberger (2007) argues that adopting the Aristotelian principle of "carving nature at its joints" through the use of taxonomies and controlled vocabularies in an increasingly chaotic and complex digital information landscape is a futile effort. Furthermore, Shirky (2005) contends that standards-based categorisation systems (including Yahoo categories, Dewey Decimal, and Library of Congress classification schemes) are best fit for physical libraries, and he goes on to state that predefined categories can only work for domains that have stable categories, limited collections and clear edges. In addition, he argues that such systems can only be functional in domains in which there are expert librarians, who are able to describe the content for an already expert user base. Both Weinberger (2007) and Shirky (2005) agree that in the digital information landscape and in domains where there exist huge collections of information objects and where users come from diverse backgrounds employing diverse numbers of terminologies in the search for information, pre-determined categorisation systems (standards-based approaches) are very likely to fail to scale and become unwieldy. Consequently, they recommend that contemporary standards-based metadata approaches should be replaced by socially-constructed ones.

In contrast, the second position suggests the limitations of socially constructed content and metadata approaches and contends that relying on a system within which structure, authority

and editorial quality are absent is likely to result in a substandard and chaotic information landscape. For example, Andrew Keen, in "The Cult of the Amateur" (2007), dismisses the notion of user generated content and uncontrolled social media as amateurish and stresses the need for control and regulation by an established authority. He argues that "what the Web 2.0 revolution is really delivering is superficial observations of the world around us rather than deep analysis, shrill opinion rather than considered judgment," and the consequence of this uncontrolled content, as Keen warns, is chaos and the disappearance of truth (Keen, 2007, p. 16). Keen argues that the unedited, uncontrolled, and unfiltered content from Web 2.0 threatens the fabric of traditional quality controlled, edited, and guarded media. In a similar vein, Michael Gorman (2005), former president of the American Library Association, stated that the Web 2.0 in general and blogs in particular, lack credibility.

The third, middle-of-the-road, position recognizes inherent limitations in both approaches and contends that finding a middle ground could rectify some of these limitations. Authors such as Gruber (2007), Smith (2008), Veltman (2001), and Wright (2007) argue that a balance should be struck between standards-based and socially constructed metadata approaches. As Gruber (2007) argues, socially constructed metadata approaches are one-dimensional; plagued with inconsistency and lack of organisation, whilst standards-based metadata approaches are forced upon users and fail to represent the users' worldviews. Gruber (2007) suggests that ontologies should be developed to capitalise on the best of both worlds. At a conceptual level, Morville (2005) and Wright (2007), contend that the two metadata approaches can productively coexist. Morville (2005, p. 139) argues that "ontologies, taxonomies, and folksonomies are not mutually exclusive." However, a closer look at Morville's arguments reveals a tendency to suggest that socially constructed metadata has its contextual place solely in the blogosphere and social media environments, and not entirely in portals and digital libraries. This is also evidenced by his choice of terminologies, such as "mob indexing" for what is otherwise termed as social classification or folksonomy, elsewhere in the metadata literature. In his book *Ambient Findability*, Morville (2005) contextualizes the theory of "Pace Layering" from Stewart Brand's *How Buildings Learn* (see Brand, 1994) and argues that "taxonomies and ontologies provide a solid semantic network that connects interface to infrastructure," whilst folksonomies are overlaid on the taxonomic metadata infrastructure providing it with the fast-moving and volatile vocabularies of users (see also Campbell & Fast, 2006; Smith, 2008). Both Wright and Morville approach the issue of folksonomies with caution, arguing against the view that folksonomies are fundamentally new ways of organizing information. However, others such as Campbell and Fast (2006) emphasize the importance of embracing Web 2.0 categorisation techniques. They contend, "We ignore 'mob indexing' at our peril; by refusing to change our methods or our approaches in the face of collaborative tagging systems, we run a serious risk of allowing information architects to become anachronisms in emerging information environments."

Chapter Four: Grounded Theory Methodology

4.1. Chapter Overview

The challenges faced by libraries in relation to the accurate description of information objects through metadata that reflects users' terminologies, and thus enhancing the findability and discoverability of information objects is indicated in Chapter One (1.1). Chapters Two and Three provided background into the research problem by reviewing existing related literature on both standards-based and socially-constructed metadata approaches. This thesis set out to explore and investigate current metadata challenges in more depth, and thus to explore alternative opportunities. Based on the researcher's interest and motivation, a broad research problem was identified. As the thesis aimed at investigating the current status of standards-based metadata approaches, identify gaps, if any, and explore alternative approaches, at the outset of this research, several research methodologies were also considered including a survey method to determine the views and perspectives of Library and Information Science professionals on metadata functions. The options of testing extant metadata theories and principles were considered. However, in view of exploring the viability of alternative metadata approaches, especially in the light of the current metadata practises versus emerging metadata approaches, extant theories that address such metadata issues are not well developed.

As a consequence, instead of the approach to testing an existing theory (deductive), theory building (inductive) was considered appropriate. It is also important to note that the research problem identification was exploratory rather than pre-formulated. A decision was then made to adopt a research methodology which allows an iterative and evolutionary process of data collection and analysis.

Hence, in accordance with the nature of reality in relation to the Web 2.0 paradigm, especially with the acceptance and inclusion of diverse perspectives on matters such as metadata description, it is argued, this thesis benefits by adopting an interpretive ontology, a social constructivist epistemology and an inductive methodology. It is indicated that, due to the possible diversity of perspectives that may be held by the potential research participants in this thesis, an inductive research methodology would be appropriate to investigate the problem at hand.

4.2. Adopting a Social Constructivist Philosophical Perspective

As recommended by Guba and Lincoln (1994), Grix (2004), Creswell (2003) and Charmaz (2006), scholarly investigation should lay its foundation on the building blocks of research. The philosophical perspectives as to whether the investigator has adopted a positivist or interpretive paradigm should also be explicitly stated at the same stage. Therefore, a clear statement of the underlying ontological and epistemological perspectives adopted in this research would help in the exploration of solutions to existing metadata challenges, and thus to develop a theory that would help to underpin future metadata practises.

The principles that underpin current standards-based metadata tend to be predicated on prior assumptions made about the terminologies that users would use to find and discover information objects (2.9.4). This can be seen for example in the use of pre-determined controlled vocabularies which contain fixed categories that are updated at wide intervals. Consequently, the terminologies are likely to be out-dated and missing context through time, and hence failing to represent users' world views accurately. Such preponderance to focus on authoritative (expert-created) approaches seem to stem from a foundationalism ontological viewpoint and tend to adopt an objectivist epistemology, thus methodologically leaving the prerogative for metadata decisions on the expert (librarian). Such a position as this, ontologically speaking, can only advocate a single solution to problems.

On the other hand, the emerging Web 2.0 paradigm (Anderson, 2007; O'Reilly, 2005) seems to indicate a move towards the acceptance of disparate points of views (multiple interpretations) and negotiated meanings regarding digital objects (Veltman, 2001; Weinberger, 2007, 2012). For example, as Shirky (2005) points out, the tension between hierarchical and collaborative categorisation boils down to a philosophical question, namely "Does the world make sense, or do we make sense of the world?" According to the author "If you believe [that] the world makes sense, then anyone who tries to make sense of the world differently from you is presenting you with a situation that needs to be reconciled formally, because if you get it wrong, you're getting it wrong about the real world", perhaps suggesting the beneficial impacts of adopting an interpretive ontological approach and a social-constructivist epistemological stance, where multiple interpretations can be considered.

In connection with this, Weinberger (2012) also asserts that knowledge is affected in one way or another by subjective experience and social interpretations and these interpretations are deeply influenced by culture, language, history, and individual circumstances. These factors, Weinberger contends, present an important challenge to the positivist view that posits every knowledge or fact can be objectively verified and thus truth can be established. Weinberger (2012) notes the problematic nature of attempting to force a single interpretation of reality. Weinberger (2012, p. 90) thus argues "you can't make sense of something outside of a context. Even something as simple as a car's turn signal can only be understood within a context that includes cars, the basics of physics, the unpredictable intentions of other drivers, the restrictions of the law, and the way left and right travels with one's body".

According to Duffy and Jonassen (1992), social constructivism posits that "meaning is imposed on the world by us, rather than being extant in the world independently of us." Social constructivism is rooted in the belief that socio-cultural and historical factors influence knowledge (Kukla, 2000). As Kukla (2000) indicates, social constructivism has its roots in the philosophical works of Marx, Mannheim, Durkheim, Merton and Kuhn. According to Duffy and Jonassen (1992, p.3) "there are many ways to structure the world, and there are many meanings or perspectives for any event or concept." This is in sharp contrast with the objectivist viewpoint that presumes that "truth and meaning reside in objects, independently of any consciousness" (Crotty, 1998). The nature of knowledge in social constructivism focuses on "individual reconstructions coalescing around consensus" thus promoting shared and negotiated meaning (Guba & Lincoln, 1989; Guba & Lincoln, 1994).

4.3. The Grounded Theory Method

The Grounded Theory Method was first developed by Barney Glaser and Anselm Strauss in 1967 (Glaser & Strauss, 1967). It was initially devised for the domains of medical sociology and health research (Charmaz, 2006; Glaser & Strauss, 1967; Strauss & Corbin, 1990). Through the years, the use of the method has slowly spread to other spheres, including information systems (Lehmann, 2010). Grounded theory is seen as a fitting method for emergent research areas, as it helps in developing conceptual foundations that are grounded in data. Allan (2007) maintains that grounded theory provides for a systematic and rigorous analysis of a phenomenon or a problem. Grounded theory incorporates proven principles and procedures, such as the use of open coding, focused coding, constant comparison, memo writing, theoretical coding and theoretical saturation (Allan, 2007; Charmaz, 2006).

The method can also be used for conceptualising real-world problems and phenomena. The main tenet of the method is the process of iterative conceptualisation, rather than description (Bryant & Charmaz, 2007a; Glaser, 2001; Glaser, 1978; Strauss & Corbin, 1998). Whilst the method is considered appropriate to collect and systematically analyse various types of data, it is well suited for the conduct of qualitative research (Charmaz, 2006). The basic tenet of the methodology is the process of developing a theory that is grounded in data through simultaneous data collection and analysis (Bryant & Charmaz, 2007a). According to Strauss and Corbin (1998, p.7) the defining characteristics of grounded theory include critical analysis, conceptual abstraction, openness to emerging ideas, and reliance on empirical data.

Similar characteristics of the methodology include avoidance of preconceived theories and pre-formulated hypothesis, as well as reflective and critical analysis of situations and contexts in any given research problem or phenomenon (Charmaz, 2006; Strauss & Corbin, 1998). The adoption of Grounded Theory Method in information science is discussed in the literature (Allan, 2007; Bryant & Charmaz, 2007a; Dunn, 2011; Lehmann, 2010; Mansourian, 2006; Nguyen, Partridge, & Edwards, 2012; Scott, 2007; Urquhart & Fernandez, 2006; Urquhart, Lehmann, & Myers, 2010).

One of the main advantages of the Grounded Theory Method is that the theory developed from data in a particular area of study (substantive theory) can help explain problems in other domains (general theory). For instance, a theory that explains users' satisfaction in library services can be employed to elucidate the phenomena in other spheres, outside of the library domain. As Allan (2007, p. 9) points out "the methodology consists of a systematic framework that, when followed, provides techniques for data analysis that are repeatable, generalise-able and more rigorous than most qualitative research methods. Grounded theory, although classified as a qualitative research method, has certain processes that are lacking in other similar methods and, hence, could be considered to be the first in a new genre of research methodologies that can be used for conceptualising underlying causal issues, rather than merely describing them."

4.4. Comparison of Grounded Theory Methodologies

Currently, there are three main approaches that can be pursued in the implementation of Grounded Theory Methodology. Each is variously referred to as after their originators and proponents, namely the Glaserian, Straussian and Charmaz's grounded theory approaches. Dunn (2011) portrays Glaser as positivistic and Charmaz as interpretivist. Charmaz (2006) also categorises both the Glaserian and Straussian approaches to grounded theory as positivist, whilst considering hers as an interpretivist. In connection with the various approaches, Dunn (2011) emphasizes the importance of choosing and consistently adhering to a particular flavour of grounded theory.

4.4.1. Glaserian Grounded Theory

The first approach, known as the Glaserian (after the originator), compels the researcher to postpone the process of literature review until such time that data analysis has been completed and a theory has been generated. Glaser is often considered as a pioneer grounded theorist and he espouses the view that the researcher should keep distance in the research process, so as not to introduce any biases and preconceived ideas.

According to Allan (2003) coding helps to differentiate the substantive from the noise in any given mass of data. The main purpose of coding is conceptualisation, which, according to Glaser (2001), is the process of abstraction of the data from time, place and people so as to be able to move beyond mere description to the identification of themes that are of value in investigating an underlying phenomenon. For Glaser (2001, p. 4) "[grounded theory] comes from data, but does not describe the data from which it emerges," and "[grounded theory] does not generate findings: it generates hypothesis about explaining the behaviour from which it is generated".

Glaser's dictum "all is data" attests that data can be collected from diverse sources: interviews, surveys and secondary sources. The method supports the use of coding, memo writing and conceptualisation. As regards to memo writing, Glaser argues that memos lead to "abstraction or ideation" (Glaser, 1978, p.83), as they provide the analyst with the freedom to reflect on his/her data.

The Glaserian Grounded Theory Method evaluates the resultant grounded theory for its fit (the categories of the theory should fit the data and not vice versa), work (provides a level of understanding that is acceptable and credible to the respondents, or some subset of them, and to the inquirer), relevance (it must deal with those constructs, core problems, and processes that have emerged in the situation) and modifiability (open to continuous change to accommodate new information that emerges or new levels of sophistication to which it is possible to rise) (Glaser, 1978; Glaser & Strauss, 1967; Guba & Lincoln, 1989).

4.4.2. Straussian Grounded Theory

The second approach, known as Straussian, came into existence when Anslem Strauss, came to hold views different from that of Glaser. Whilst Glaser advocated strict adherence to the original tenets of the method, as contained in "The Discovery of Grounded Theory" (Glaser

and Strauss, 1967), Strauss, along with his colleague Juliet Corbin, argued that the method should evolve in accordance with pragmatic situations (Strauss & Corbin, 1998). Strauss and Corbin argued that the method should be evolving as a tool of research (Strauss & Corbin, 1990). They admit the importance of recognising bias that may creep by the researcher's preconceptions (Strauss & Corbin, 1998).

In terms of objectivity of the researcher, Strauss and Corbin (1998) acknowledge that the researcher cannot possibly go into the problem without any pre-conceptions, however they advise that the researcher should try to stay as objective as possible during any interactions with participants.

Glaser was also strongly opposed to Strauss and Corbin's detailed procedures for data analysis, as contained in their "Basics of Qualitative Research Techniques and Procedures for Developing Grounded Theory" (1990), going as far as advising the authors to "either re-write their book or re-name the new method" (Glaser, 1992). Glaser is convinced that there is just one Grounded Theory Method, one that he calls the 'true grounded theory', staunchly opposing all other flavours (Glaser, 2001). These differences led to the first split in Grounded Theory Methodology.

4.4.3. Charmaz's Constructivist Grounded Theory

The third flavour is attributed to Kathy Charmaz (2006) who argues that both Glaser and Strauss were wrong-footed in being positivists in their treatment of the researcher as a distant and objective observer during data collection and analysis. Charmaz's approach is known as the Constructivist Grounded Theory Method as it adheres to a constructivist philosophical approach, wherein both the researcher and participants mutually co-construct meaning during data collection and analysis. According to Charmaz (2006, p.131), the positivist approach to grounded theory lends itself to the objectivist and deterministic approach to research, where it considers the existence of a single interpretation to reality. For Charmaz, both Glaserian and Straussian approaches to grounded theory treat the researcher as an objective observer. In contrast, Glaser (2002) contends that Charmaz's notion of co-construction of interpretations between researcher and participant biases the results. Charmaz (2006) argues both the Glaserian and Straussian grounded theory approaches emanate from an objectivist stance, thus giving way to the emergence of this third approach, the Constructivist Grounded Theory Method.

As Charmaz (2006) and Mills et al. (2006) emphasise, in Constructivist Grounded Theory Method, the interaction between the investigator and participants in interviews cannot be neutral. Mills et al. (2006) argue that through active engagement during the interview process, ideas are raised, discussed and knowledge is mutually constructed. According to this view, the researcher and the participants co-construct data, in a process known as data generation.

Charmaz (2006) and Mills et al. (2006), advocate non-hierarchical intimacy, reciprocity, open interchange of ideas and negotiation (including consensus on the location and time of interview) between the researcher and participants. The researcher also has the opportunity to

express and reflect upon his/her viewpoints and perspectives (Mills et al., 2006), in a way similarly to what happens during other conventional conversations and academic discussions. By acting thus, the interviewer has the opportunity to voice his view points and perspectives while allowing the voices of interviewees to be heard. Furthermore, the method should be allowed to evolve without losing its main tenets, namely, simultaneous data collection, avoidance of pre-formulated hypothesis, systematic coding, constant comparisons, theoretical sampling and theoretical saturation (Bryant & Charmaz, 2007a, 2007b). In addition, unlike Glaser, Charmaz argues the voice of the research participant should be part and parcel of the various stages of analysis including in the final writing up process.

Finally, it is important to note that Charmaz (2006) argues against Glaser's original notion of not conducting a literature review. Charmaz contends that as researchers have already been exposed to the extant theories, concepts and hypotheses in their fields of study, it is impractical to assume that staying away from doing a literature review helps in steering clear of such pre-conception. Charmaz has no problem with extant theories in the literature, as long as the researcher is aware of them and takes care not to start from them. Charmaz still agrees with the view that the theory should emerge from the data. However, her constructivist flavour of grounded theory suggests that she believes that theory emerges from an active engagement between the researcher and participants during the interviews and other data collection stages.

4.5. Rationale for Choice of the Constructivist Grounded Theory Method

The patterns of behaviour, views and perspectives of research participants is considered the core component, hence, grounded theory is well suited for the study of these attributes. As Lehmann (2010) acknowledges, grounded theory is an appropriate method for research in information systems, as the domain deals with several overarching components, including technology, data, procedures, and people. Allan (2007) also asserts that grounded theory is a systematic and rigorous method for researching information systems. He details how its various procedures, such as the identification of Open Codes, Categories and Core Categories, can be used when conceptualising real-world problems in information science research and can help in generating theory that explains patterns in behaviour, users' satisfaction or other relevant research issues. It has been acknowledged that the method is especially pertinent in areas where there is scarcity of theoretical foundations. The sub-category of information systems research that deals with digital libraries is one such domain, as it is one in which the generation and use of theories has been scant to date (Andersen & Skouvig, 2006; Floridi, 2009; Hjørland, 2000; Lehmann, 2010). Andersen and Skouvig (2006, p. 318) assert that "for knowledge organisation to uphold significance recognisable by society, it needs to engage in and be informed by theories and understandings that locate and analyse society and its historically developed forms of organisation". There is, therefore, a pressing need for developing theories. It is hoped that, in the context of this thesis, a constructivist grounded theory method would inductively generate Categories and Core Categories, which in turn would help the development of a theory.

Noting the scant usage of grounded theory in LIS research, Mansourian (2006) indicates that the simultaneous data collection, iterative conceptualisation and rigorous interplay with data are some of the most important tenets of the method which are beneficial for LIS research. Due to the importance of the emergence of concepts from the ground up, as Mansourian indicates, it is relevant that sufficient time is allocated, suggesting thus the relevance of the method for long-term research projects, such as PhD (Mansourian, 2006). Mansourian (2006), however, does not indicate which of the three grounded theory approaches best fits for LIS research.

As indicated in 4.2, this thesis takes an interpretive ontology and social constructivist epistemology. Taking into account, disparity in the nature of digital libraries, their collections and the varying user needs, the potential existence of diversity of views among librarians, LIS researchers, metadata experts and library users with respect of the issues of involving users in metadata creation, from the three approaches to grounded theory, Charmaz's constructivist approach was considered fitting. Adopting a constructivist epistemological approach and Grounded Theory Method, it is argued, affords the researcher flexibility and rigour to gather views and opinions, through interactive and iterative in-depth interviews. It also allows the researcher to analyse and interpret the perspectives of participants' through identification of Open Codes, Categories and Core Categories from the data collected. Finally, the method is expected to help to develop a theory that overarches the Core Categories derived from the data collected.

4.6. The Constructivist Grounded Theory Process

4.6.1. Literature Review

Reviewing extant literature helps the researcher to highlight the conceptual background within the substantive area under study (Chapter Two & Chapter Three). It also helps in the final portion of the research discussion and relating the research questions to what has already been investigated elsewhere by other researchers (Charmaz, 2006, p.168). The conventional wisdom, received from grounded theorists, is to approach the problem with an open mind, but not with an empty mind (Charmaz, 2006; Glaser, 1978; Strauss & Corbin, 1990). This is agreed upon by most proponents of the method and the technical term used by them is theoretical sensitivity (Glaser, 1978).

Charmaz (2006) recommends that the researcher embark on research with some tangible problem at hand, along with any pre-conceived ideas and knowledge about the problem. Charmaz argues that the issue of theoretical sensitivity to the research problem at hand is inescapable, which, according to her, is something to be encouraged. "Give earlier works their due", Charmaz (2006, p.166) re-iterates. She believes reviewing the literature helps to identify gaps in extant works, place the research in context, refine, extend or revise existing theories, and to "weave the discussion" in the light of earlier works. Thus, she emphasises the importance of reviewing the literature with a critical mind. Charmaz (2006, p.165) notes the various routes researchers take in terms of the timeline of a literature review including whether it is necessary to postpone it until the completion of the grounded theory analysis.

Attending to the importance of flexibility, Charmaz seems to leave the decision of the timeline to the researcher. In light of this, the review of related literature was conducted (see Chapters Two and Three).

4.6.2. Data Collection Techniques

The Grounded Theory Methodology allows simultaneous data collection and analysis (Charmaz, 2006; Guba & Lincoln, 1989). The methodology allows data to be collected using field notes, interviews, historical documents and government records. However, the data thus collected should be weighed in terms of relevancy, quality and quantity (Charmaz, 2006, p. 16). One of the most widely used data collection techniques in grounded theory is intensive interviewing. This technique allows the researcher to have an in-depth exploration of a topic, with the interviewer's active engagement, and interpretation of the interviewee's responses. As Charmaz (2006, p. 26) describes it, "an [intensive] interview goes beneath the surface of ordinary conversation and examines earlier events, views and feelings afresh".

4.6.3. Theoretical Sampling

According to Strauss and Corbin (1998, p. 201), theoretical sampling is the process of identifying categories that emerge repeatedly from the data and aids concentration on an examination of the properties and dimensions of the core category. As Charmaz (2006, p.96) explains, theoretical sampling helps in saturating the Core Categories with additional empirical data, until no new properties emerge (see also 4.6.7).

4.6.4. Data Analysis through Coding

The Grounded Theory Method proceeds in an iterative interchange of data collections and analyses. Memo writing is an important part of this process (Charmaz, 2006; Strauss & Corbin, 1998). The various coding strategies and the memos written during the analysis phases are essential for identifying Open Codes, Categories and Core Categories as well as for developing a theory subsequently. Coding is an essential step in a grounded theory data analysis. As Charmaz (2006, p.43) defines it, coding is the process of labelling a line, sentence or paragraph of interview transcripts or any other piece of data (such as segment of audio tape and video record) with a short and precise name.

Although the different flavours of grounded theory prescribe different stages of coding, following Charmaz (2006), three stages of coding have been adopted in this thesis: open coding, focused coding and theoretical coding.

During Focused Coding, the researcher generates the "bones of analysis" which will then be integrated and assembled at the stage of theoretical coding, which is crucial for identifying emergent Core Categories for further analysis and subsequent theory development (Charmaz, 2006).

Coding follows from a detailed analysis of the data obtained from interview transcripts and questionnaires. Whilst it is the researcher's prerogative as to whether to assign new labels or utilise the exact expressions employed by the participants, commonly referred to as 'in vivo'

codes in the literature (Charmaz, 2006, p.55; Strauss & Corbin, 1998, p.105), the labels/terms should be able to describe the underlying data and also evoke meanings and actions. It is worth noting that 'in vivo' codes were first used by Glaser and Strauss (1967). Charmaz (2006) and Allan (2007) also recommend the use of gerund verbs as they help to identify dimensions and provide richness for analysis. This is to say the term 'describing' (verb) is preferred to 'description' (noun); or 'leading' (verb) to 'leader' (noun). As Charmaz (2006, p.49) asserts gerunds carry with them "a strong sense of action and sequence while also helping to remain focused on participants' responses and contextual meanings".

Overall, Charmaz (2006, p.49) identifies the following key issues that need to be kept in mind during coding: remain open minded (whilst also recognising the difference between an open-mind and an empty-mind, as openness here refers to the importance of allowing concepts to emerge and not forcing preconceived concepts onto the data), stay close to the data; use simple, short and precise codes; preserve actions (use of gerunds), ensure constant comparisons between responses and concepts; and move quickly through the data; capture/condense meanings into "compelling codes [that] capture the phenomenon and grab the [attention of] the reader" (Charmaz, 2006, p. 48). As the Grounded Theory Method is an intensely iterative process, the researcher should modify codes; re-word them with catchy/grabbing phrases through such iterative process. As Charmaz advises, "make your codes fit the data, rather than forcing the data to fit your codes" (p.49).

The three stages of coding, namely open coding, focused coding and theoretical coding, are discussed below.

4.6.4.1. Open Coding

The first phase is known as open coding (initial coding). It refers to the analytic stage in which concepts, their properties and dimensions are identified (Strauss & Corbin, 1998, p.101). Charmaz refers to this stage as initial coding (2006, p.47), however, the nomenclature 'open coding' characterises the process better as it indicates that the identified codes are grounded on the data and that the researcher remains open to any new codes as well as any in vivo words. Allan (2007) advises that, during open coding, the researcher should keep asking: "What is this data a study of? What code does this incident indicate? What is actually happening in this data? This will continually remind the researcher of the original research intentions and aids him/her to stay in focus without getting lost amongst masses of data". He further recommends: "Don't analyse too much data at one go, in other words carefully examine the transcripts. Don't be totally biased with preconceived concepts. Stop and write memos in between. Don't lose track of your research topic. Ignore data that is not pertinent to the research topic" (Allan, 2007). The process of Open Coding results in a flat collection of Open Codes.

4.6.4.2. Focused Coding

Focused Coding, as the name indicates, is a directed coding procedure and it involves the decision to bring similar codes into a nested category of codes (Charmaz, 2006). Focused Coding is an abstraction of the codes generated during the Open Coding stage. This stage of

data analysis results in comparing, linking, refining and abstracting open codes into a hierarchy of codes called categories. As Charmaz (2006, p.70) notes coding in general “routes [one’s] work in an analytic direction” and Open Coding is the main stage of Constructivist Grounded Theory analysis where the data starts to make sense. The resultant codes of Focused Coding are referred to by Charmaz as Focused Codes whilst other constructivist grounded theorists, such as Glaser and Strauss refer to them as Categories. For the analysis and discussions in this thesis, the term Categories is used to distinguish them from open codes.

4.6.4.3. Theoretical Coding

In order to further abstract and conceptualise the categories, a third stage of coding called Theoretical Coding helps to integrate the various Core Categories. Whilst Focused Coding facilitates the organisation of the Open Codes into Categories, Theoretical Coding, enables the emergence of the Core Categories. Charmaz (2006, p.63) notes that theoretical coding helps to “hone your work with a sharp analytical edge”. The use of memos (containing reflections on Core Categories) and constant comparison between Core Categories are instrumental for Theoretical Coding. During each of these refinement and saturation processes, the analysis moves from mere description to conceptualisation.

4.6.5. Writing Memos

In Grounded Theory Method, memo-writing is an important step in the conceptualisation of data. Memos serve the researcher as analytic tools (Charmaz, 2006, p.72); helping him to pause and reflect on the data collection procedure and on the data collected. They also provide insight and are also helpful in deliberating on why a certain participant holds a particular point of view. According to Charmaz (2006, p.80) memos should be kept informal and can be written at either the early stage of data collection (early memos) or at the later stages of data analysis (advanced memos).

4.6.6. Theoretical Saturation

Theoretical sampling (determining what type of data to collect next) and theoretical saturation (terminating data collection) are two important processes of Constructivist Grounded Theory. Theoretical saturation is said to have been achieved when the core categories that have emerged from the research process are saturated (developed) with adequate data to the extent that the incorporation of new data provides no additional insight. Saturation is the stage at which the core categories, identified during the analysis, are supported through relevant and rigorous data and thus the various properties of the categories are established in great detail (Charmaz, 2006).

According to Charmaz (2006, p. 100) initial sampling helps in determining where to start data collection. It is, nonetheless, important to note that this decision is also partially pragmatic, since, although there is always the possibility that issues other than those covered by the current research exist in the wide world, one has to stop somewhere. Charmaz (2006, p. 114) also adds that theoretical saturation is a subjective exercise and that the Constructivist

Grounded Theory Method, being an interpretive approach, acknowledges both the importance and limitations of such subjectivity.

4.6.7. The Process of Selection of Research Participants

One of the features of the Grounded Theory Method is that the number of respondents (sample size) cannot be determined in advance. Instead, a procedure called theoretical sampling is employed. It is a technique that guides data collection as the study progresses, on the basis of the Categories that have already emerged from an analysis of the data that has been collected at a previous stage. Once a problem has been identified and an initial location for the study has been selected, initial data gathering may start at any place selected by the researcher and participants. As the initially collected data is analysed, some preliminary Categories will begin to emerge. These Categories will then guide the next phase of data collection. Such theoretical sampling continues iteratively until such time that theoretical saturation is reached. The latter is the stage at which additional data stops providing new insights about the Categories (Coleman & O'Connor, 2007).

According to Razavi and Iverson (2006, p.461), in the Grounded Theory Method “informants chosen for interviewing must be expert participants, with rich, extensive prior experience with the phenomenon, in order to be able to provide the researcher with a valid account of their experience.” However, everything that comes to add value to the research problem or phenomena is deemed relevant. As mentioned earlier, the first series of data collection and its subsequent analysis will serve as a guide to the next stage of data collection. This methodological approach is in accord with the constructivism paradigm. Guba & Lincoln (1989, p. 180) for example advise: “As the design proceeds, the constructivist continuously seeks to refine and extend the design - to help it unfold. As each sample is selected, each datum recorded, and each element of the joint construction devised, the design itself can become more focused. As the constructivist enquirer becomes better acquainted with what is salient, the sample becomes more directed; the data analysis more directed the construction more definitive”.

4.6.8. Criteria for Evaluating a Constructivist Grounded Theory

In a constructivist approach, the purpose of the final write-up does not seek to discover “truth” and does not provide a generalisation either (Guba & Lincoln, 1989, p. 180). According to Charmaz (2006, p.182), grounded theory studies should be evaluated, firstly, for its credibility, which refers to the rigor in which the identification and development of open codes, categories, and core categories are supported by empirical data collected by the researcher. In connection with this, Charmaz (2006, p.182) asks whether “the data is sufficient to merit your claims?” The second criterion is originality which directly shows the contribution of the core categories to extend or challenge existing practices. According to Charmaz, resonance is the third criterion, which is to what extent the findings make sense to the people involved (affected) by the findings. In other words, the participants should make sense of the core categories and the theoretical rendering that resulted from the analysis of their data. The fourth and final criterion of Constructivist Grounded Theory, according to

Charmaz (2006), is usefulness. Usefulness answers how the core categories, and subsequently the theory that emerged from the data, is relevant to inform actual practises and should contribute to existing knowledge (Charmaz, 2006, p.183).

4.7. Conclusions

By describing the tendency of the Web 2.0 paradigm to adhering to multiple perspectives and viewpoints and contextualising it to the issues of library metadata, this chapter provided the rationale for the choice of an interpretive ontology, social constructivist epistemology and an inductive methodology. The chapter described the main tenets of the Grounded Theory Method as an inductive research methodology. It is indicated that the methodology affords the researcher the ability to explore emerging concepts through simultaneous data collection and analysis, iterative conceptualisation using such techniques as coding, memo writing, and theoretical saturation. The chapter compared the three approaches to the Grounded Theory Methodology: Glaserian, Straussian and Charmaz's Constructivist Grounded Theory. Rationale for using the method for library and information research was reviewed and, pursuant to the novelty of socially-constructed metadata approaches, the Constructivist Grounded Theory Method is considered as a fitting method. It is suggested that being an inductive method, this research does not test an existing hypothesis, but it develops novel ones through the research process itself. Thus, this research takes an inductive approach where Categories and Core Categories are developed from empirical data collected using intensive interviews. In addition, the Constructivist Grounded Theory approach affords the researcher the opportunity to be part of the research process, through proactive mutual co-construction and reflexivity, during data collection and data analysis. The following chapter presents the detailed research design and procedures adopted in this thesis.

Chapter Five: Research Design using the Constructivist Grounded Theory Method

5.1. Overview of Research Design

This thesis is informed by empirical data gathered in the course of three separate but interrelated studies:

- **Study One** (Initial or Proof-of-Concept Study with expert users): The study was designed with the aim of exploring the broad issues of standards-based and socially-constructed metadata approaches, so as to identify the main concerns and thus to delineate the scope of this research. Using intensive interviewing technique (see Appendix-2.1), the perspectives of a total of 11 Library and Information Science academics and postgraduate students were explored and discussed. As this was a Proof-of-Concept study, it does not use the complete Constructivist Grounded Theory coding procedure, instead used the initial broad categories to set the semi-structured interview questions for Study Two;
- **Study Two** (Main Study with LIS Professionals, Librarians and Metadata Experts): the study was designed with the aim of identifying Open Codes, Categories and Core Categories using the Constructivist Grounded Theory Method (Open Coding, Focused Coding and Theoretical Coding) (see Appendix-2.2);
- **Study Three** (Saturation of Study Two with library users): the study was designed with the aim of saturating the results of Study Two with the perspectives of library users using the Constructivist Grounded Theory Method (Open Coding and Focused Coding) (see Appendix-2.3).

The first, **Study One** (Chapter Six), was conducted as a preliminary, proof-of-concept, exploration of broad issues on standards-based metadata approaches and the implication of the Web 2.0 paradigm on metadata, in a very open-ended manner. Since Grounded Theory Method permits neither a hypothesis to test nor initial set of well-defined and structured sets of research questions, in this thesis, a decision was made to conduct preliminary stage of data collection so that the analysis from this stage identified the key issues in relation to the broad area of investigation. In light of this purpose, after identifying the general research problem, this research set out to investigate the views and experiences of Library and Information Science (LIS) academics and postgraduates regarding standards-based and socially-constructed metadata approaches. This group of participants were considered as expert users.

The selection of interviewees for Study One was essentially purposive. The researcher was affiliated with the European Commission-funded International Masters in the Digital Library Learning program. The program attracts a diverse mix of international students who have worked in libraries and selected to study under the program. A total of eleven in-depth

interviews were conducted, with two lecturers, one PhD student and eight MSc students in Digital Library Learning (DILL), all of whom were associated, in one capacity or another, with the EU-funded, Erasmus Mundus program at Tallinn University, Estonia. Countries of origin of interviewees included Bangladesh, China, Denmark, Estonia, Ethiopia, Italy, Turkey, Uganda, Venezuela and Zimbabwe. The results of the Proof-of-Concept study are presented in narrative discussion (Chapter Six). Whilst, this is not a representative sample, the responses obtained from the participants is very rich and informative.

Based on Study One, **Study Two** was designed with a more specific set of semi-structured questions and in-depth interviews were conducted with twenty-one LIS researchers, librarians, and metadata experts. Study Two uses the full Constructivist Grounded Theory Method. Results and discussions of Study Two are presented in Chapters Seven to Twelve. Since both Studies One and Two are based on the perspectives of Library and Information Science (LIS) professionals, it is essential to explore the perspectives of library users (see Appendix-2.3) in view of their experiences in using libraries and its resources, in particular that of the library catalogue. Thus Study Three was conducted (see Chapter Thirteen).

5.2. The Intensive Interviewing Process

All the three studies used intensive interviewing as the data collection method. The intensive interviewing technique was chosen, in order to enable the interviewer to ask for more detail, delve into an issue, go back and forth among important points and request more explanation (Charmaz, 2006). The interview processes were open-ended, conversational, and mutually constructed, hence it ensures that the required depth, richness and rigour is acquired. Semi-structured interview questionnaires were prepared (see Appendix-2).

Since the objective of the study was to iteratively identify, saturate and develop emerging concepts, the selection of participants was purposive. In particular, prospective interviewees were systematically identified through prior contacts or an identification of interviewees' work experiences and research interests. In accordance with the Constructivist Grounded Theory Methodology, the choice of all potential interviewees was not pre-determined. Instead, the process was iterative and evolutionary. For example, the first sets of interviews were transcribed, reflected upon through memo writing, and then used as a basis for categorizing, discovering, selecting, informing and getting the consent for subsequent sets of interviewees. The number and category of interviewees is shown in Table 5.1.

Study No.	Interviewees		
	Study Total Number	Educational Level	Sub-total
Study One: Proof-of-Concept Exploration	11	LIS MSc Students	8
		LIS PhD Student	1
		LIS Lecturers	2
Study Two: LIS Professionals	21	Librarians	10
		LIS Researchers	5
		LIS Lecturers	2
		Metadata Consultants	4
Study Three: Library Users	25	Under-graduate Students (BSc., BA)	4
		Post-graduate Students (MA, MSc)	4
		Post-graduate Students (PhD)	6
		Lecturers (Other than LIS)	11
Total Number of Interviewees			57

Table 5.1: Categories of Interviewees

Unlike descriptive statistical research methods, the Constructivist Grounded Theory Methodology favours fewer participants, but necessitates more detailed and intensive interviews. Therefore, the number of interviewees in this research is relatively fewer (a total of 57 for all three studies) when compared to other kinds of research methods, such as surveys. This is partially due to the Constructivist Grounded Theory Methodology's focus on identifying and developing concepts on the basis of a few, but intensive, data collection endeavours, rather than aiming at representation and generalisation that forms the essence of other research approaches. Following Charmaz (2006), this research followed the procedures and techniques depicted in the Research Design diagram, shown in Figure 5.1.

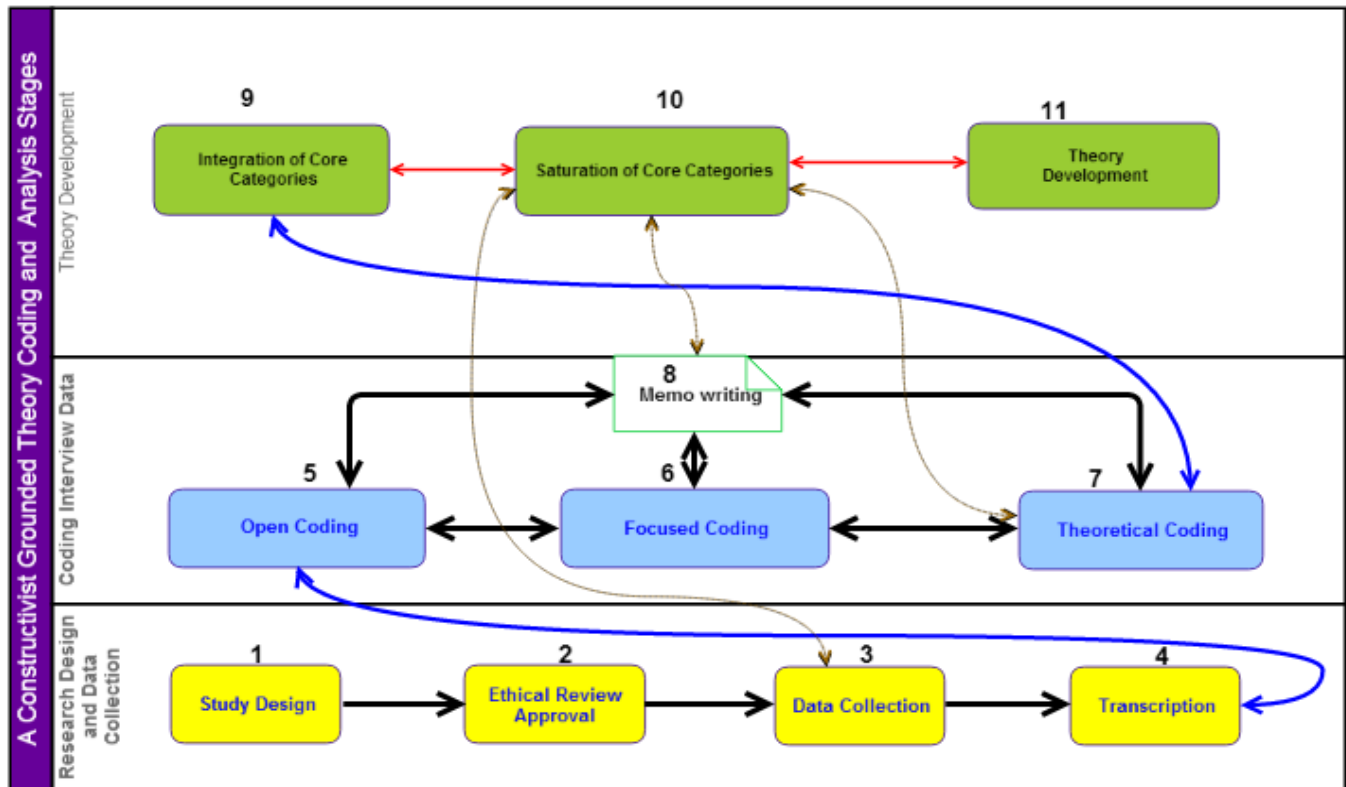


Figure 5.1: Research Design for Study Two and Study Three (based on Charmaz, 2006)

Full definitions of the procedures portrayed in the Research Design Diagram (Figure 5.1), which form constituent elements of the Constructivist Grounded Theory Methodology, are to be found in 4.6. As is evident from Figure 5.1, (note the double edged arrows), the Constructivist Grounded Theory Methodology is an iterative, and non-linear, evolutionary process.

5.3. Ethical Review Approval for the Three Studies

In compliance with the University of Portsmouth's requirements that proper ethical mechanisms are established and strictly followed during research, all three studies conducted in this thesis have been reviewed and approved. For Study One, the application for a full ethical review and approval was submitted to the Ethical Review Committee of the Faculty of Creative and Cultural Industries, University of Portsmouth, on February 2nd, 2011 and was approved on February 17th, 2011 (note that the formal letter of approval was written later in on June 8th, 2011, with an Ethical Approval Number: FO:01/11-0046, see Appendix-1.1). The application included a filled-in "Full Ethical Review Application for Researchers" form, participants' information sheets, consent forms, email correspondences with initial contacts with prospective interviewees, and interview questionnaires. The Ethical Review application for Study Two was submitted on December 19th, 2011 and was approved on January 10th, 2012, with an Ethical Review Number: FO: 01/12-0054 (see Appendix-1.2). Similarly, an ethical review application form for Study Three was filed on October 11th, 2012 to the CCI Faculty Ethics Committee and was approved on December 5th, 2012 under the ethical review number FO:11-12-0068 (see Appendix-1.3).

5.4. Use of Computer Assisted Qualitative Data Analysis Software (CAQDAS):

NVivo 9

The use of software for qualitative data analysis is mainly for efficiency purposes. It enables the organisation of interview data and also facilitates the various stages of coding, memo writing and integration of the various emergent concepts and categories. As Atherton and Elsmore (2007) point out, the careful use of software provides efficient data handling and organisation capabilities for researchers. Advocating the use of Computer Assisted Qualitative Data Analysis software (CAQDAS), Welsh (2002) notes that software can provide an audit trail of the data analysis process. It is important to note that, unlike quantitative data analysis software, such as SPSS, CAQDAS offers very little help in terms of generating automated data analysis procedures. In the context of Constructivist Grounded Theory data analysis, the use of CAQDAS is limited to data organisation and retrieval, coding segments of interview data with labels, creating a hierarchy of codes, and, finally, eases the process of memo writing. To put it another way, the use of CAQDAS provides efficiency and flexibility.

There are a number of CAQDAS applications available in the market. NVivo 9 is one such application, designed with Grounded Theory data analysis in mind (Welsh, 2002). As Atherton and Elsmore (2007) advise, the choice and use of software for qualitative data analysis should match the underlying methodological and philosophical assumptions of the specific research at hand. Hence, NVivo version 9 was chosen to support data analysis in this study (see Figure 5.2).

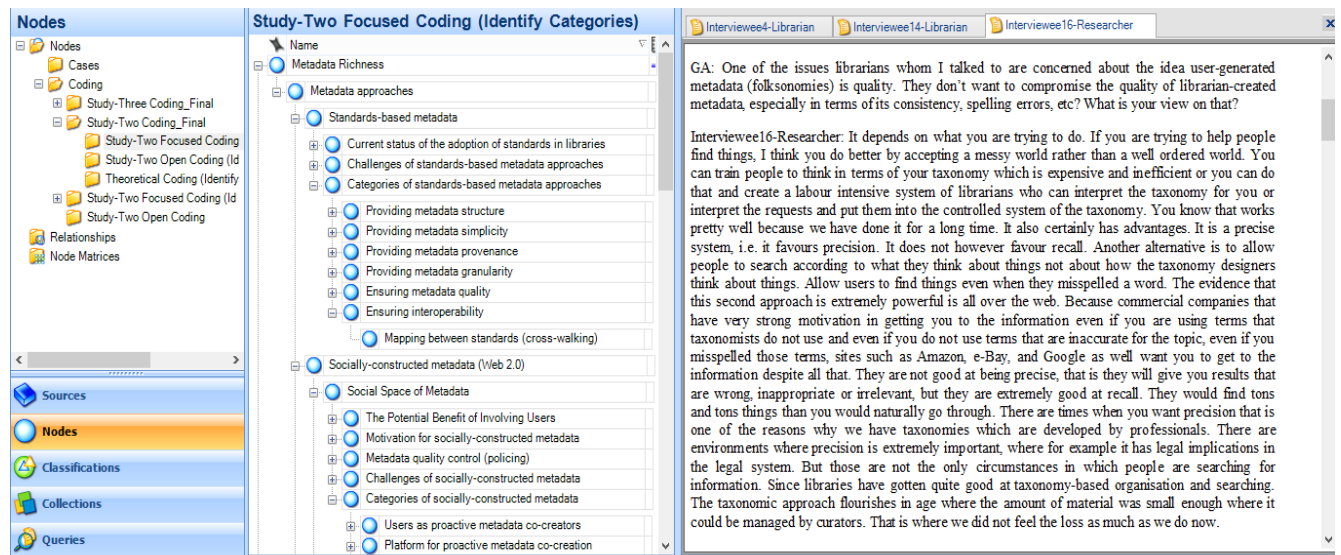


Figure 5.2: NVivo 9 Interface

The software enables coding interview data with labels, called Open Codes, creating relationships between Open Codes, placing Open Codes into Categories, and writing memos about Categories and Core Categories (Figure 5.2). The software also offers the capability to easily navigate among documents, which otherwise would have been unwieldy.

5.5. Data Analysis Method and Workflow Using Three-Stages of Constructivist Grounded Theory Method

Following Charmaz (2006), the Constructivist Grounded Theory Method adopted for analysing the data collected during the intensive interviews, had three coding stages; Open Coding (identifying open codes), Focused Coding (identifying emerging categories), and Theoretical Coding (identifying emerging core categories and developing theory) (see Figure 5.3).

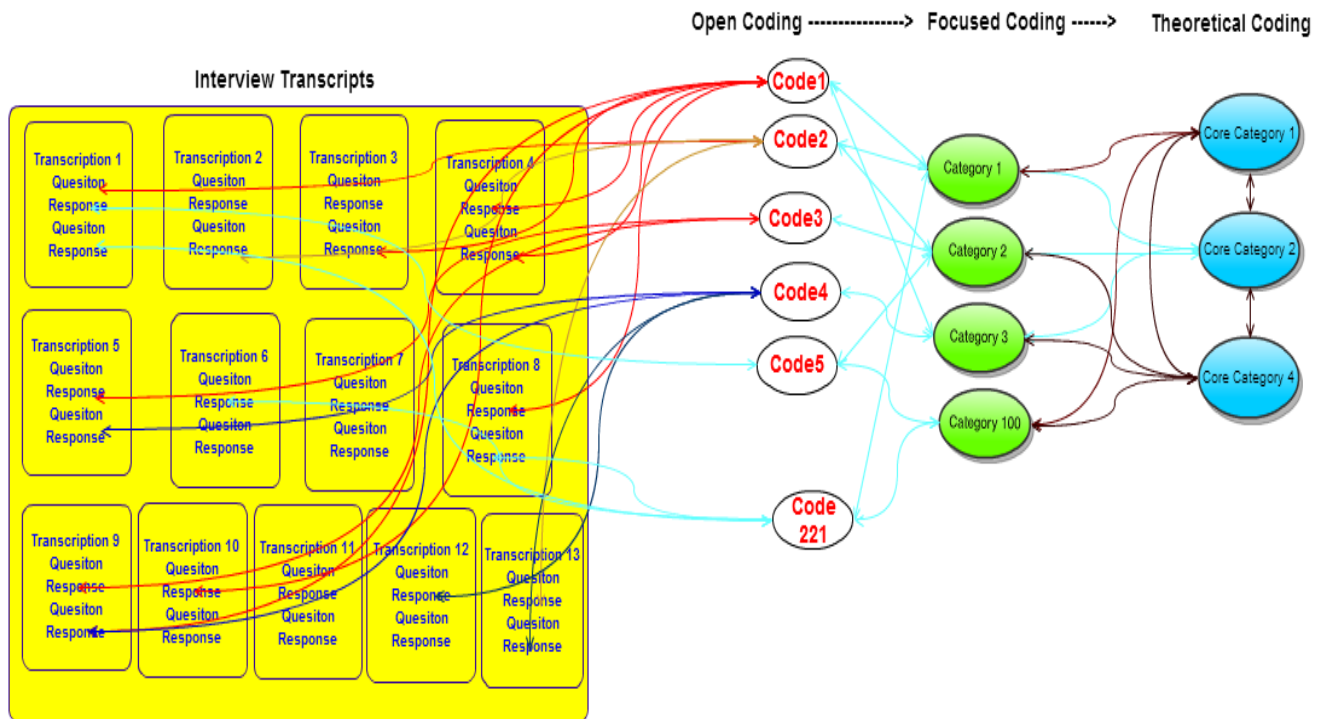


Figure 5.3: Three Stages of Constructivist Grounded Theory Coding (based on Charmaz, 2006)

In conformance with the Constructivist Grounded Theory Methodology, data collection and analysis procedures were implemented iteratively (Figure 5.3). The various Codes, Categories and Core Categories evolved slowly, in an emergent process. The Constructivist Grounded Theory coding was used for studies Two and Three. Since the focus of the Theoretical Coding is conceptualisation rather than description, the emergence of a Core Category from Categories does not necessarily imply a hierarchical (top-down) relationship.

5.6. Conclusions

This chapter discussed the processes and procedures adopted in this thesis in the investigation of the perspectives of research participants (interviewees) on issues pertaining to descriptive metadata, mainly in relation to standards-based and socially-constructed metadata approaches. Following Charmaz (2006), a Constructivist Grounded Theory Methodology was adopted and three stages of analysis, namely, Open Coding, Focused Coding and Theoretical Coding, were designed and executed. It was indicated that the study design had fully met the requirements of the Ethical Review and Approval process of the University of Portsmouth.

The chapter also discussed the rationale behind the choice of an in-depth interviewing technique and the use of NVivo as a computer assisted qualitative data analysis software platform. In addition, the chapter discussed the use of three data analysis stages, namely Open, Focused and Theoretical Coding.

The following chapter (Chapter Six) presents the Proof-of-Concept (Study One) and presents discussions and narratives with regard to the perspectives of LIS postgraduates and academics. Subsequent chapters (Chapters Seven to Twelve) provide details of the Constructivist Grounded Theory analyses and elaborate on the Categories and Core Categories. More precisely, Chapter Seven and Chapter Eight are devoted to detailed analyses of Open Codes and Categories emerging from the analysis of the perspectives of LIS professionals on standards-based and socially-constructed metadata approaches, respectively. Thereafter, the results of Theoretical Coding, i.e. Core Categories, will be examined in depth in Chapters Nine, Ten, Eleven, and Twelve.

Chapter Six: Study One (Proof-of-Concept Study) Preliminary Exploration of the Perspectives of LIS Professionals on Contemporary Metadata Approaches

6.1. Overview

Whilst it is possible in the Grounded Theory Method to directly delve into the main research study without conducting a pilot study, this thesis designed and executed a preliminary Proof-of-Concept exploration into the broad topic of descriptive metadata. The aim of the study was to delineate the overarching concepts which would henceforth help to ring-fence the scope of the subsequent (main) study. For the Proof-of-Concept, the intensive interviewing data collection technique was adopted and it consisted of open-ended questions on broader issues of descriptive metadata including both standards-based and socially-constructed metadata approaches.

Prior to each interview, introductory contacts were made, via e-mail, in order to obtain the consent of each interviewee as well as to reach a bilateral consensus as to the timing and venue of the meeting. According to Charmaz (2006), “the in-depth nature of an intensive interview fosters eliciting each interviewee's interpretation of their experience.” Thus, the actual interviews were made purposefully informal, in order to encourage dynamic participation on the part of interviewees in the ensuing discussions, which were expected to constitute a significant portion of the meeting.

The results from the analysis of this study suggest the anachronistic state of the library Online Public Access Catalogues (OPAC) (6.3), the lack of integration of the Web 2.0 paradigm in existing standards-based metadata systems (6.5 & 6.6), the need to re-conceptualize current metadata approaches and the importance of developing a theory (6.11) that is aimed at addressing the needs of library users. To this end, the analysis of interviews also suggests the centrality of representing users' perspectives in the creation and utilisation of metadata. The results from this analysis will help to guide the research by providing concepts and issues which can serve as a spring board for further exploration.

6.2. Library as a Place

The interviewees, who are themselves librarians by profession or being trained to be librarians, disclosed that they rarely go to the library physically. This was mainly a consequence of their ability to access information resources, in the form of electronic information services such as library databases and e-journals, from their study and work places. Some respondents hold the view that the library is not a place that is particularly important for their work. One interviewee (Study1-Interviewee1) in particular went as far as stating that they had never gone to a library during the past two years. Two other interviewees (Study1-Interviewee3 & Study1-Interviewee4) asserted that they mostly relied on remote access to electronic sources, subject-librarian assistance, and information gathered through their network of researchers and academic experts in order to keep abreast of

developments in their particular areas of research, as well as to receive recommendations regarding pertinent articles and other publications. Consequently they had no particular interest in physically visiting the library.

In a similar vein, one interviewee (Study1-Interviewee2) disclosed that they considered the time spent in browsing books and journals in physical libraries as a waste of time and stated, “If I get access to a database, the first thing that I would ask is whether it has a remote access feature, using a password, that would enable me to use it either at home or at the university; thereby, hopefully, eliminating the need to go to the library.” This attitude was shared by another interviewee, who indicated that they hardly ever went to the library. The interviewee's exact words were “I don't even know where the library at [this university] is exactly located” (Study1-Interviewee8).

However, the interview population needs to be born in mind. As Library and Information Science academics and postgraduate students, the interviewees may be more aware than others of digital library facilities available. Whilst the issue of whether physical academic libraries are relevant is not a matter for consideration in this chapter, what the findings of this study reveal is a pattern of preference on the part of “busy” university academicians and postgraduate students for conserving time through seamless-remote connectivity to an array of library resources, accessible at their convenience. Such behaviour clearly demonstrates the importance of efforts aimed at making digital library collections more visible, easily discoverable/findable, and usable. To this end, the necessity for instituting richer metadata becomes more important. The perspectives of library users on the library as a physical and virtual place is also discussed in Chapter Thirteen (see 13.2).

6.3. Obsolescence of Online Public Access Catalogue (OPAC)

Most interviewees were unanimous in stating that they found the OPAC outdated, especially in comparison with most currently popular web search engines. In support of this view they point out, for example, that most OPACs do not seem to have an alternative spelling option. The absence of this and similar, seemingly simple, features makes the OPAC less user-friendly. In addition, most OPACs do not have features that enable users to rate, comment, review, or share resources with others, a serious shortcoming, in view of most users' needs.

Most interviewees asserted that they rarely use the library's OPAC. One interviewee noted, “I am not quite sure if I am using the OPAC quite extensively as a catalogue [...] because somehow it seems to me that I've created my own personal library on my computer. I even like to call it my desktop library” (Study1-Interviewee4). Likewise, another interviewee expressed that “the OPACs have a big problem in that they were instituted just to replace card catalogues [and] it didn't evolve adequately enough with developments within the world of the Web. Hence the catalogues and other library instruments always tend to be a little behind when compared to search engines” (Study1-Interviewee1).

However, one interviewee (Study1-Interviewee2) described the OPAC as the “biggest innovation for libraries that has ever happened,” although further suggesting that libraries are changing too slowly in trying to cope with users' novel needs and expectations. The

interviewee noted in particular that current cataloguing systems follow “refined ways of cataloguing [oriented to assuring metadata quality]. Mistakes, for instance, are not allowed in catalogue searches. But, in reality, users make mistakes.” According to this interviewee, the search engine Google reflects reality better than those libraries that stick to strict cataloguing rules.

In a discussion of how the OPAC is performing in the current information landscape, one interviewee remarked that libraries are trying to catch up, but “when things are no longer cool anymore, then librarians start doing them, [such as] implementing the Web 2.0 paradigm in the OPAC” (Study1-Interviewee8). As a result of not incorporating Web 2.0 in time, the interviewee noted that most present OPACs lack interactivity.

Thus as a whole, as the responses from interviewees revealed, their preferences are for Google-like search interfaces, in which a user is allowed to search with any combination of keywords. This is in agreement with a report prepared by Calhoun (2006, p. 5), who suggested that a “large and growing number of students and scholars routinely bypass library catalogues in favour of other discovery tools, and the catalogue represents a shrinking proportion of the universe of scholarly information.” This is partly because, as Choy (2011) argues, and interviewees also agree, that library catalogues are becoming inconvenient as they require users to accurately remember and submit bibliographic details such as Author, Title, ISBN, and other details during the information retrieval process. Bates (1989), Borgman (1996), Buchanan and McKay (2011), Evans (2009), and Fifarek (2007) also discuss the disconnect between search techniques including the choice of terminologies by users and what is expected of them by designers of library systems, emphasizing the importance of incorporating the vocabulary of users into library metadata systems.

6.4. Preference for Searching Versus Browsing

In response to the question of whether interviewees preferred searching, browsing, or both, most replied that they liked to start by searching with a search engine or in a database, and follow that with browsing through the retrieved hits for relevant items. One interviewee (Study1-Interviewee2) pointed out the limitations of browsing, noting the problems associated with scalability as the number of pertinent documents increases. This interviewee's search strategy was a walk-through, from the use of keywords in a search box, to refinement of retrievals, until satisfactory results were obtained. The interviewee also stated that they mostly used subject keywords rather than authors' names or the titles of documents. Describing experiences in physical libraries, the interviewee noted that “a librarian myself, I sometimes fail to locate a book on the shelves.” Regarding the use of keyword searching, one other interviewee (Study1-Interviewee7) said, “What I prefer is just full text search [using] key words”.

Another interviewee (Study1-Interviewee1) discussed their personal information discovery technique as one that employs “navigation trails.” This approach, the interviewee elaborated, starts with a keyword search (which was criticised for sometimes becoming too noisy, though it was still the better option available), then scanning/reading one or more of the full-text documents, then looking for more sources from the list of references. This process, the

interviewee remarked, leads one to the most pertinent authors and other publications cited in that document. These trails, the interviewee explained, allow one to “step from one publication to another.” The interviewee also noted that the reference list, compiled by the author(s), is the core part of this “trail,” emphasizing that such trails, not only provided links to related works that support each other, but also led to works that constituted contending views. The downside of this trail, the interviewee pointed out, is that “you need to be a good librarian to navigate through these trails of references. It also requires patience. But once you get your way through, it helps in the discovery of related works whilst enabling one to build up his/her knowledge.” One interviewee (Study1-Interviewee1) stated that they began an information search from Google and then cross-checked for trustworthiness of the sources.

Overall, the responses from interviews indicate a preponderance by interviewees of beginning an information search from popular web search engines rather than the library catalogue. It must be borne in mind that, in addition to preferences for searching with one's own keywords, the responses also highlighted the need for getting access to the full-text of publications, rather than merely the bibliographic citation. One of the interviewees (Study1-Interviewee5) remarked thus, “Regarding access to content, the first [thing] I consider is [whether] it is full text or not, full access or partial. This is the most important consideration [for me]. Because I do not lack bibliographic information, I need to read the full text as partial text is not enough for me” (see also 13.4.4). This indicates the interviewee's preference for immediate access to full-text publications, such as PDF files, during their search. It was also indicated that most OPACs, currently, do not allow users to use a combination of unstructured keywords. In addition, the interviewees prefer conducting searches rather than browsing, which has an important implication on the way metadata is created and utilised.

6.5. Standards-based Metadata Approaches

All 11 interviewees acknowledged the proliferation of standards. Some even repeated a modern adage, “The good thing about standards is that there are so many you can choose from.” Under these circumstances, making the correct choice, when selecting a standard, as well as ensuring interoperability between digital libraries is bound to be a daunting task, although interviewees are not currently involved in these decisions. One interviewee expressed the opinion that “libraries should base their [selection] decisions on the type of resources [they have] and the subjects they are describing.” During the discussion, it became apparent that, for this interviewee, interoperability was a much sought after issue, even though it was a complicated one. Nevertheless, the same interviewee conceded that established standards such as Dublin Core, Machine Readable Cataloguing (MARC), and Library of Congress Subject Headings should still be maintained by libraries, as they are the basis for fulfilling various library functions.

Interviewees have indicated the complexity of MARC vis-à-vis the simplicity of Dublin Core, noting that the relative ease of use comes at the expense of metadata richness. It was also pointed out that some standards, such as MARC, have been unnecessarily stretched too wide, in order to enable them to accommodate certain other genres of information objects. In particular, one interviewee had this to say:

“We now have a novel information landscape, but a standard that is anachronistic, whilst most OPACs lack interactivity, as they are essentially static. Currently, libraries are so fixated on such archaic standards that they look a little outdated when compared to search engines, such as Google. MARC is a complicated but rich, standard. In contrast, Dublin Core is simple but lacks descriptiveness and richness” (Study1-Interviewee1).

It has been pointed out that the existence of several metadata standards as well as the variety of purposes for which they are used, along with linguistic differences among users, is bound to create interoperability problems. An interviewee (Study1-Interviewee5) remarked, “In [my country, there are] about 80 universities; [yet] we don't have a union catalogue because of the multiplicity of standards currently in use.” He also noted that, even whilst using the same standard, different libraries might interpret it differently so that the same book can be given non-identical call numbers in different libraries. Another interviewee (Study1-Interviewee9) observed that “libraries in China adopt and modify international standards according to their needs.” The interviewee went on to mention that they sometimes find it difficult to understand the similarities in standards used in Europe or America when compared to those used in China, simply because the Chinese versions have been extensively modified, in order to accommodate local needs. Part of the problem, this interviewee noted, can be attributed to the nature of the Chinese language. Similar concerns were voiced by one interviewee (Study1-Interviewee1) who remarked that, at present not only is the number of standards proliferating but that the same is true regarding the solutions to metadata interoperability.

Another interviewee (Study1-Interviewee3) recalled their experience of helping in the design of a database that complied with a Soviet-era standard, known as VINITI and administered by the All-Union Institute for Scientific and Technical Information. After a lot of work that involved describing database objects in accordance with this standard had been accomplished, it was later stipulated that the database should yet be compliant with the European standard. This caused a huge amount of technical rework. Elaborating on the dissimilarities among standards, the said interviewee observed that even Dublin Core and the IEEE Learning Object Metadata (LOM) standard had differences that resulted in incompatibility issues. The same interviewee characterised LOM as being a complicated standard, since filling out all the elements used for describing a learning object was cumbersome, based on recollections of experiences gathered whilst using LOM in another project, this time whilst working with Nordic colleagues. In addition to its complexity, the interviewee recollected several difficulties arising from the absence of some fields that were required for describing their learning objects.

The interviews highlighted the challenges arising from the presence of a plethora of metadata standards. Such moves are likely to put increasing pressure on interoperability among disparate information systems (Chan & Zeng, 2006; Dekkers, 2002; Dempsey, 2004; Haslhofer & Klas, 2010; Miller, 2000; Nagamori & Sugimoto, 2006; Nilsson, 2010; Nilsson et al., 2006; Ouksel & Sheth, 1999; Rothenberg, 2008; Veltman, 2001; Zeng & Chan, 2006) (see also 7.3 & 7.4).

6.6. Sluggish Adoption of the Web 2.0 Paradigm for Metadata Creation and Utilisation in Libraries

In connection with the adoption of the Web 2.0 paradigm in libraries, one interviewee (Study1-Interviewee3) reflected on how some Web 2.0 applications come and go. The interviewee stressed the need for questioning the purpose served by employing a specific technology before adopting its use in library functions. The same interviewee cited the example of Second Life and how libraries had adopted it just to be part of Second Life, and hence, librarians started creating their own avatars without addressing the rationale behind its use. Elaborating on experience, the interviewee observed that, currently, the use of Second Life in libraries has diminished. Interviewee (Study1-Interviewee3) also added that they do not foresee Web 2.0 replacing the old systems of information organisation. However, many interviewees felt that libraries should adopt the Web 2.0 paradigm for metadata creation and utilisation. Some of the interviewees were, however, wary of the lack of control and structure in Web 2.0 technologies (e.g. tagging) and applications (e.g. Flickr). When it came to the usage of standards-based and socially constructed metadata approaches, there was a consensus among the interviewees' responses that existing classification systems and the new collaborative approaches of tagging can be utilised together and should not be considered as incompatible methods. Thus, the interviewees advised that libraries need not necessarily discontinue current metadata functions or abandon current classification schemes, as doing so would mean removing those users who utilise them.

In answering the question of how interviewees perceived the adoption of the Web 2.0 paradigm for library metadata creation, as compared to the standards-based approach, one interviewee (Study1-Interviewee1) replied that they were not particularly a "big fan of the folksonomy and social web stuff." The same interviewee expressed a belief in the superiority of a more structured system and the roles played by librarians in creating such structured systems. According to the interviewee, they found it difficult to abandon a library-oriented view of the world, which, the person added, might be biased. From the conversation, it became evident that, the interviewee did not totally dismiss folksonomies, as they later stated the probable usefulness of employing tagging as a complement to current metadata usage. The interviewee concluded by observing that the social approach of tagging could be utilised side-by-side with structured metadata created by librarians. Nevertheless, the interviewee advised that the two types of metadata should be kept separate and be maintained at different levels, in such a manner that a user could switch from one approach to the other, as the need arises.

In a discussion focusing on categorisation and classification systems, one interviewee (Study1-Interviewee2) gave details of the techniques used for organizing personal music and games collections on their own computer. For organizing the songs, the person relied on common, widely used tags, such as the name of the singer, the title of the song, the name of the group who plays it, the year it was issued, and so on, which, the interviewee stated, did not pose any particular difficulties. The interviewee, however, mentioned a problem when trying to make sense of the differences among the various music genres defined by publishers. This is because, as the interviewee noted, music genres are very subjective,

similar to an art collection. The problem, the interviewee indicated, arose from the fact that most genres reflected the requirements of American music radio stations; hence, some genres were irrelevant for the interviewee's purposes. The interviewee pointed out "genres such as: Adult contemporary, Album oriented rock, and Soft rock are not meaningful for me. You may know what hard rock is, but soft rock is not that comprehensible." As a result, the interviewee had had to reorganize the genres and rename some of them, based on information obtained from Wikipedia, which, according to the interviewee, had a system that better reflected the person's needs.

Another interviewee (Study1-Interviewee1) agreed that tagging is a good thing but expressed concerns regarding the absence of control, citing synonym and homonym ambiguities as problems afflicting tagging systems. For example, the interviewee selected the term Torino, which may refer to a city, a football club, or the car company in Italy, thereby, creating ambiguities. According to the same interviewee, tags, by themselves, are not enough as they should be used with the more structured and organised metadata created by librarians. For this interviewee, the claim that tags do not have any structure and control does not make a lot of sense. The interviewee noted that current tagging systems, such as LibraryThing, do seem to have some form of a structure as, for example, the description of a book consists of standard metadata elements such as title, author, and publisher.

Another interviewee (Study1-Interviewee9) pointed out that Web 2.0 technologies such as social tagging bring similar information together and provide information that is very practical for and widely known by the lay community. They observed that libraries traditionally offered highly formalised knowledge, such as text books, but for daily usage, the interviewee said, they preferred using web search engines. A deeper grasp of the concepts and terms employed in a given topic of interest offered the opportunity to coin keywords, for use in these search engines. Another interviewee (Study1-Interviewee4) remarked that they took a "softer" approach toward the Web 2.0 paradigm, in that they saw it as a useful approach for augmenting library functionality. Elaborating, the same interviewee stated, "If we are talking [about the fact] that the library is meant for users, I think we should not only consider taking users' needs and wants into account, but they should also be involved when systems intended to serve are designed and developed" (Study1-Interviewee4).

Similarly, one interviewee (Study1-Interviewee9) also expressed the view that Web 2.0 is important for practical applications. The interviewee cited, as an example, the world of gardening, about which, the interviewee said, they would be interested in discovering information that is expressed in terminologies of day-to-day usage and laymen's terms, rather than in scholastic ones, as found in books from libraries. One good example, the interviewee observed, was the use of standard metric units in formal text books, whilst most of these metrics are rarely employed in everyday language.

Most interviewees indicated the importance of putting in place some control and structure in the adoption of Web 2.0 for library metadata. According to one interviewee (Study1-Interviewee1), even Wikipedia, arguably considered the freest and most democratic approach, has some basic rules and structure underlying it. The interviewee maintained that

what makes Wikipedia different is the fact that it has no scope. Nor does it have a rule stipulating “let us create these numbers of pages for these numbers of topics.” Wikipedians, the interviewee asserted, act according to the following maxim, “Let us put everything out. If you want to add voice, you can do so. If you want to make the voice louder, you can do that too. If the voice stays [hushed], because nobody is contributing, that is ok. It is so heterogeneous.”

As the interviewee (Study1-Interviewee1) explained, Wikipedia has aims, to at least present a neutral point of view. This, according to the interviewee, is one of Wikipedia's few rules. Additionally, one should be able to convey other, even contrary points of views. Furthermore, Wikipedia also watches out for offensive language and imposes some writing style. Hence, according to the interviewee, overall, Wikipedia “is not completely devoid of rules and there is always a structure.” As further expressed by the interviewee, the case of Wikipedia demonstrates that a user-driven contribution can be supported within a structure, albeit a minimal one.

Overall, as the responses of the interviewees and a review of the literature demonstrate, the implication of Web 2.0 technologies for library metadata functions merits a closer examination (see Chapter Eight). Contrary to the above, another interviewee (Study1-Interviewee2) pointed out that the current usage of Web 2.0 technologies in libraries has not been well thought out. As an example, the interviewee singled out the current usage of Facebook by libraries, whereby both users and librarians make postings. However, as the number of postings grows, searching and/or browsing through them becomes increasingly cumbersome. The problem, the interviewee opined, was that Facebook had not been designed to serve as a search engine. In addition, at present, most of libraries’ postings on Facebook are limited to such basic facts as opening hours and location. In the same vein, the interviewees remarked about the use of LibraryThing, Delicious and other social tagging applications. It is important to note that the adoption of Web 2.0 in libraries it not well integrated and combined with current metadata functions. Without changing current metadata functions, libraries simply buy-in or implement one or more Web 2.0 applications. It is thus imperative that the adoption of the Web 2.0 paradigm for libraries is re-conceptualised. Of paramount importance is the need for libraries to avoid “adopting technology for technology's sake”.

6.7. Sluggish Adoption of Linked Data in Libraries

Another question posed to interviewees was the role of the Semantic Web for metadata functions in libraries. After acknowledging being aware of the visions of Semantic Web technologies, many of the interviewees confessed to being unaware of any Semantic Web application appropriate for use in digital libraries. One interviewee (Study1-Interviewee1) expressed a belief that the Semantic Web would offer solutions to the problems that they had previously highlighted during the interview, including cataloguing and storage of information in a manner that can be searched semantically. Within this context, the interviewee foresaw “Linked Data” becoming an effective method of metadata representation, at the most discrete and atomic data level. The interviewee continued, saying that “you just describe at the

meaning level and then create associations. Such meaning will lie not with the object itself but on the context of the associations. The OPAC should be more robust and allow natural language searches.” Similarly, another interviewee (Study1-Interviewee2) characterised the Semantic Web as “a system built on meaningful relationships between topics. The world is modelled in words. And then you can see the relationships between the words and what kinds of relationships there are. It is a good technology but I don't see libraries using it. And I don't currently see a proper Semantic Web.”

Another interviewee (Study1-Interviewee3) stated that they had heard a lot about the Semantic Web but hadn't come across any real instance of its application in libraries. Another interviewee (Study1-Interviewee7) described the Semantic Web as a system of creating structured metadata to describe information resources and other objects. The interviewee added “to tell you frankly I [have] never used any Semantic Web application. But I think from what they say, it would be quite a difficult task”. One of the interviewees (Study1-Interviewee8) portrayed the Semantic Web as “an awesome idea” but expressed a reservation as to whether they would need the agents proposed by Berners-Lee, Hendler, & Lassil (Berners-Lee et al., 2001), who had envisaged a role for semantic agents in facilitating such tasks as coffee making and fixing appointments. Most of the promises of the Semantic Web, according to a recent review by Powell, Black, and Collins (2011), “Have yet to be fulfilled.” It is perhaps worth mentioning here that Thomas Gruber, who is mainly known for his oft-cited definition of the term ontology (i.e. “an explicit specification of a conceptualisation” (Gruber, 1993), has for so long been advocating the role of semantic agents not from an artificial intelligence point of view but from ontology-based metadata specification of concepts, objects, and its relationships. It is also important to note here that Gruber is also behind the development of the Siri software, a semantically intelligent virtual personal assistant, which uses Semantic Web technologies.

Most interviewees acknowledged the potential use of Semantic Web technologies for information organisation and access. In connection with this, one interviewee (Study1-Interviewee5) called attention to the technical complexity of Semantic Web technologies. Linked Data, the same interviewee observed, is an important part of the Semantic Web. This interviewee also predicted that the Semantic Web would be more popular in the coming five years. The interviewee's prediction doesn't seem to be farfetched, given the fact that these technologies have promised to provide concrete solutions for metadata representation and utilisation as early as 2003 (Day, 2003b) and subsequently in 2008 (Rothenberg, 2008). Another interviewee (Study1-Interviewee9) concurred, stating that they understood the Semantic Web as a concept, but stressed that the technologies associated with it should be taught in computer science departments. The interviewee foresaw applications ensuing from a wider deployment of these technologies, for which development and acquisition of new skills, by both students and users, is a prerequisite.

6.8. Bridging Cultural Differences

All 11 interviewees stated that metadata should reflect linguistic and cultural diversity, adding, however, that ensuring this would be a huge undertaking. Elaborating the difficulties

thereby involved, one interviewee (Study1-Interviewee5) cited the example of the place name Ephesus, which has historical, religious, and cultural significance based on Christian and Islamic perspectives. To be sure, this is just a simple example of the problems arising from homonyms in terminologies employed in metadata systems. The interviewee deemed it essential that, in metadata systems, varying interpretations are reflected appropriately and that all diversity in views is presented accurately. Expanding further, the same interviewee explained, “When one associates the term Ephesus with a beer brand in Turkey, another one may relate it to a place mentioned in the Qur’an, whilst somebody else might link it to Christianity [such as Paul’s Epistle to Ephesians]. Hence we should put several cross references there.”

But the same interviewee conceded that metadata should not confuse users, maintaining, “You can describe term[s] using as many as 100 keywords, but unless you have order, there is no meaning.” Another interviewee (Study1-Interviewee9) pointed out that, in the Chinese language, there are many ways of describing the same object, based on the context (a.k.a. emotion). They singled out, as an example, the term spring (water) in a mountain, claiming that there are several (perhaps as many as nine) ways of describing it. The interviewee asserted that when current metadata systems, such as Dublin Core, are used, only one of the contexts (emotions) can be represented and the remaining eight are lost. Even worse, the person maintained, is the situation whereby should a single character be missing in the metadata description, then the whole meaning is significantly affected.

It has been shown that semantics is a contentious issue. Being culturally bound, anything can be interpreted or understood in a variety of ways, depending on cultural and social contexts. One interviewee (Study1-Interviewee6) remarked that even the seemingly trivial question of how people use their phones may result in a wide variety of responses. The interviewee asserted that the naming and labelling of terms as metadata should be predicated on such issues as the scope of the services that are aimed to be accomplished by employing the metadata. According to one of the interviewees, important questions to be considered include, “Is it universal? Does it serve international students? Is it intended to serve people residing in a particular locality? Once you get the answers to these questions right, then you can tell how you should be able to correctly define the meanings and the labels associated” (Study1-Interviewee6).

A different interviewee (Study1-Interviewee4) pointed out that libraries have quite a diverse category of users. The interviewee maintained that whilst some users utilise formal language, others may employ informal wording for describing their information needs. The interviewee continued, elaborating:

“A person may come to the library and state that he wants the book he was reading yesterday. Another might say he wants a yellow book, or a book with the red dots on the left side of the title. For me, as a librarian, the latter requires some kind of interpretation. As for a layman, his information needs will not be as sophisticated as for a professor teaching in a class room. Nevertheless, that is still a need. So for me, a

user-centred library means that formal academic language is not a special (exceptional) one” (Study1-Interviewee4).

The need to incorporate formal and informal languages of users in library information organisation functions coheres with arguments put forward by Veltman (2001) where he argues that cultural diversity should be reflected in metadata descriptions.

6.9. Collaboration is Essential but Missing

According to one interviewee (Study1-Interviewee3), the discipline of information organisation, including metadata, suffers from a lack of collaboration among professionals, just as with many other fields of human endeavour. The interviewee cited the case of Dublin Core and IEEE's Learning Object Metadata (LOM), identifying them as two incompatible standards. The interviewee realised that using LOM was too complicated, stating that “unless you are a librarian or unless you have special interest in the discipline, as a teacher, there is much to be done to add all these metadata for your learning objects.” The same interviewee even claimed that it was as if there were a conspiracy to make things more complicated.

Another interviewee (Study1-Interviewee2) stressed the need for libraries to provide richer descriptions of library collections, recommending that librarians should collaborate worldwide in order to be able to do so. According to this interviewee, librarians should be permitted, by their institutions, to catalogue collections of other institutions and vice versa, instead of relying on metadata records from proprietary companies. The interviewee pointed out that “the way Asians describe Asian art is quite different from the way a Westerner does.” The same interviewee dwelt on the need for collaboration among librarians across countries and institutions. The interviewee even went to the extent of recommending:

“Libraries should crowdsource their cataloguing. Hence, libraries should pay their staff to help catalogue the world. A library in Australia could be cataloguing a library collection in Tallinn. The staff in Tallinn could catalogue a book in New Zealand ... A good model would be to collaborate with other librarians all over the world” (Study1-Interviewee2).

As stressed by the interviewees, collaboration among metadata stakeholders is highly important. One of the Web 2.0 concepts interviewees mentioned is metadata crowdsourcing. Crowdsourcing, a phrase coined by Jeff Howe in 2006, refers to a new model of outsourcing in which an open call is made using Web 2.0 platforms and a set of tasks that are accomplished over the network by crowds of people (Adams & Ramos, 2010) (for detailed discussion on metadata crowd sourcing see 8.4.1.5). In light of metadata richness, crowdsourcing has huge implications for libraries. It can be used to solicit metadata for specialised and niche collections that cannot be described by hired librarians for reasons that librarians do not know enough about these objects. This may include adding metadata to books in different languages, labelling old maps with place names, tagging photographs, deciphering obscure letters, and so on. These can be done for both profit and altruistic motivations. In regard to this, in his recent book, *Cognitive Surplus: Creativity and Generosity in a Connected Age* (2010), Clay Shirky postulates how people without profit

motivations could collaborate in creating value that, aggregately, would exceed what could be achieved in institutional contexts. In his earlier book, Shirky (2005) cites Wikipedia as a typical example of the result of collaboration (see 3.10). Similarly, the generation of metadata could, and should, be approached from such cognitive surplus approaches.

6.10. Standards-based Versus Socially-Constructed Metadata Approaches: A Philosophical Difference

One interviewee (Study1-Interviewee1) stated that the issue of using standardised approaches (such as hierarchies and categories) versus socially-constructed metadata approaches is more of a philosophical nature. In support of this, the interviewee cited the literary work known as the Divine Comedy, by Dante, where the organisation of the poem reflects the theoretical (philosophical) framework of medieval European society. According to the interviewee, the work is a complete summary of all the medieval beliefs and church teachings extant then. Furthermore, the division of the poems is well thought out, each category having 33 divisions, which, along with the introduction, brings the total number of categories to 100. They then contrasted this with the Dewey Decimal Classification system. The same interviewee remarked that both Dewey and Dante represented cultural frameworks of their societies and that they were correct in their own way. The interviewee noted that the situation now is entirely different “because there are too many traditions altogether and we don't believe any more in a rigid, [monolithic], structure. We [do] believe in change.”

As noted by almost all interviewees, there exists some bias in current classification systems and standards. This is in agreement with Van House's (2005) assertion that “classification systems and categories carry their history within them, including the politics of the time and place in which they are created, and the participants in the decision making.”

Hence, it is important that information organisation systems reflect the diversity of users' perspectives and interpretations of information objects that have been deposited in digital libraries. To this end, as Van House (2005) advises, the philosophical assumptions that underlie standards or categorisation systems should not “valorise” one view whilst disparaging others. In their oft-cited book, *Sorting Things Out: Classification and Its Consequences*, Bowker and Star (1999) contend that classification systems exhibit inescapable limitations to adequately represent the perspectives and interpretations of users. To this end, the decisions to choose one metadata approach over another or the choice of standards or perhaps the absence thereof should be supported with sound theoretical foundation.

6.11. The Need for Theoretical Foundations

As one interviewee (Study1-Interviewee3) pointed out, the importance of answering the “why question” in using a particular technology should be correctly addressed. “If there are new developments, new tools, whatever it is, first we should have some kind of understanding about it and then we should think and [answer] why we are using it? It shouldn't be like, oh... yea, Second Life, it is hot topic, let us use it. But why is it good for libraries? Why it is better than what we already had?” The interviewee (Study1-Interviewee3) contextualised the

Second Life example to metadata and emphasised the importance of answering the question “why do we need a new metadata standard on top of existing ones?” Researching and understanding the underlying reasons and rationalising our decisions, they underscored, is crucial.

Whilst addressing the issue of the availability and use of theoretical foundations in LIS in support of services and functions, another interviewee recollected a discussion they had with colleagues from Nordic and Baltic countries. As the interviewee recalled, the argument focused on whether LIS is a theory or model-based discipline. Whilst acknowledging the existence of models, such as Tom Wilson's model for information search behaviour and Carol Kuhlthau's information search process model, the interviewee pointed out that LIS, as it stands at present, borrows rather than develops theories. Although remarking that this might not necessarily be such a bad thing, the interviewee emphatically stated that such theories be properly adapted and contextualised.

This assessment is in agreement with the opinions of several researchers (Andersen & Skouvig, 2006; Day, 2010; Floridi, 2000; Lehmann, 2010). Writing on the same topic, Floridi (2000, p. 37) recommends that “the development of LIS should not rely on some borrowed, pre-packaged theory.” Similarly, Day (2010) contends that “we [in LIS] are struck by the poverty of foundational theories.” It is also important that one takes into account socio-technical issues (Lagoze, 2010). Citing Van House, Bishop and Buttenfield (2003), Lagoze (2010, p. 6) argues that, “the socio-technical perspective is particularly appropriate for [the] analysis of information technologies (e.g. digital libraries) because of the manner in which the creation and exchange of information is so deeply embedded in almost all human activities.” The importance of theoretical frameworks in LIS has also been raised by Virkus et al. (2009). Svenonius (2001, p.ix) also argues that “instant electronic access to digital information is the single most distinguishing attribute of the information age. The elaborate retrieval mechanisms that support such access are a product of technology. But technology alone is not enough. The effectiveness of a system for accessing information is a direct function of the intelligence put into organising it. Just as the practical science of engineering is undergirded by theoretical physics, so to the design of systems for organising information rests on an intellectual foundation”.

6.12. Conclusions

As remarked by interviewees, provided that they are able to access information resources at their convenience; they find little or no reason to go to the library in person. This behavioural pattern highlights the importance of making electronic library resources discoverable, perhaps with the provision of richer metadata. The idea of providing richer metadata is considered an important issue worth further exploration. The concept of user terminologies and the importance of re-conceptualising current metadata approaches were also highlighted.

The interviewees characterised the current state of the OPAC as being anachronistic, especially when compared with contemporary search engines. It has also been reported in the literature that there is a disconnect between the designs of OPAC systems and the search behaviour of users. Interoperability problems among disparate digital libraries, arising from

the proliferation of metadata standards, have also been brought to light by several interviewees in the domain. It has also been suggested that existing standards-based metadata systems be re-evaluated, especially in light of socially constructed metadata approaches.

Whilst, it has been remarked by interviewees that these two metadata approaches do not stand in opposition to each other, interviewees have recommended that, libraries should embrace the Web 2.0 paradigm strategically, rather than adopting it for the sake of the technology. There is consensus among interviewees that, in as far as they would like to see the adoption of the Semantic Web, there is still a lack of implementation for library metadata. The interviewees anticipated that, once Linked Data is implemented and adopted widely, the Semantic Web has a lot to offer for metadata representation and exchange.

Conceptually, in this chapter, interviewees have argued that there is a need to reconceptualise current metadata systems in light of changing user needs, expectations, and evolving vocabularies. Hence, the creation and utilisation of metadata should be underpinned by sound theoretical frameworks. In accordance with the above, from the narrative description and discussions, a range of conceptual themes emerged, including the obsolescence of the current OPAC, the sluggish adoption of the Web 2.0 paradigm and Linked Data, the importance of collaboration, metadata diversity and the need for theoretical foundation to support metadata functions.

These themes along with their constructs provide an important avenue to explore with detailed and rigorous studies. Thus, the main study (Study Two) will provide a detailed investigation using empirical data from professionals in the domain. Whilst this chapter showed the importance of underpinning metadata decisions with conceptual foundations, the concepts and principles need yet to be explored, identified and developed, thus the main study will aim at identifying and developing these concepts.

Methodically, the Proof-of-Concept study showed the workability of the intensive interviewing technique, especially it evidenced the importance of focusing on a few but experienced and knowledgeable interviewees in the area under study in order to get in-depth and rigorous responses for the otherwise open-ended questions asked. This was in accord with, Charmaz (2006), Mills, Bonner and Francis (2006, p.10), who advise to follow a non-hierarchical intimate, reciprocal, open interchange of ideas and negotiation (including agreeing on the location and time of interview) with the interviewees. In this study, the method adopted offered both the researcher and interviewees an opportunity to reflect on their viewpoints and perspectives. This was also in accord with the interpretive ontology and social constructivist epistemological stance taken in this thesis (see Chapter 4.2). The intensive interviewing technique resulted in relevant discussions supported by open-ended questions which helped to cater the conversations towards the experiences and areas of expertise of the interviewees'. As shown in the Proof-of-Concept study, the intensive interviewing technique proved to be working, thus providing support that it could work well for subsequent studies.

Since this was the Proof-of-Concept study, providing a descriptive narrative of the interviewees' perspectives was found sufficient instead of using the complete Constructivist Grounded Theory Method's coding procedures. The Proof-of-Concept shows the relevancy of exploring the perspectives of library and information science (LIS) professionals on the optimality of standards-based versus socially-constructed metadata approaches.

The conceptual themes that emerged (such as the obsolescence of the OPAC in relation to web search engines, the preference of users for searching versus browsing, as well as the prolificacy of extant metadata standards resulting in interoperability challenges, the role ascribed to Web 2.0 and Semantic Web technologies in libraries, and the importance of enriching information objects with metadata that better conveys the various perspectives of users) helped to develop more detailed questions for Study Two. In other words, the narrative discussions obtained from the Proof-of-Concept in Study One helped in framing and scoping Study Two.

In addition, the discussions in the Proof-of-Concept showed the importance exploring in depth the roles that socially-constructed metadata approaches could play. It was thus found essential to identify the perspectives of LIS professionals including librarians, lecturers, researchers, and metadata consultants on the various facets of socially-constructed metadata approaches, including tagging, reviews, ratings, and recommendations. Additional specific areas of exploration on socially-constructed metadata approaches included metadata quality concerns, user motivation, and the potential roles of Linked Data (see Appendix 2.2).

Chapter Seven: Study Two - Open and Focused Coding of the Perspectives of LIS Professionals on Standards-based Metadata Approaches

7.1. Overview

The processes of research design, data collection and data analysis procedures chosen for Study Two have already been described in Chapter Five. This section presents the findings of the data analysis stage, resulting mainly from the two stages of coding: Open Coding and Focused Coding. It also traces how the **Open Codes** were identified during Open Coding of interview data and the **Categories** that emerged from **Focused Coding**, thus providing evidence for, as well as serving as a link as to how subsequent analyses and discussions are grounded in empirical data.

Whilst the interview questions formulated for data collection in Study One had been relatively few in number and essentially open-ended (see Appendix-2.1), the ones for Study Two were more numerous and relatively more specific and focused. Thus a matrix of semi-structured interview questions and corresponding constructs (see Appendix-2.2) were devised, in order to ensure the exploration of the various research facets.

In accordance with the research design (see 5.1 through 5.4), interview data was collected from 21 LIS professionals including librarians, researchers, lecturers and metadata consultants. As per the Constructivist Grounded Theory Method, the data collection was done in a slow and evolving process, thus the 21 interviews were collected over the course of one year (January 2012 to December 2012). The interview data for Study Two is the main empirical evidence for this thesis, for which the interview questions were derived from the preliminary (Proof-of-Concept) study (see Chapter Six for narrative discussions; see also Appendix-2.2 for a Matrix of Semi-Structured Interview Questions).

For Study Two, the interviewees represent a diverse mix of personalities, each having different experiences, authority and expertise in the domain of Library and Information Science, including heads (directors) of metadata and bibliographic services at world-renowned international, national and academic libraries, notable, well-experienced and published researchers, internationally recognised metadata consultants who are also involved in international metadata standards development, experienced faculty members and practising librarians based at various institutions across the world. Prior to the selection of an interviewee, a prospect's publications as well as on-going research projects were reviewed so as to gather as rich and diverse a view as possible from him/her and also invoke interest in participation in the present study. Interviewees' places of work included the British Library (3), Library of Congress (1), Harvard University (1), University of Portsmouth (3), University of Loughborough (1), Kings College London (1), University of South Australia (1), University of Bologna (1), University of Parma (1), University of Zimbabwe (1), University of North Texas (1), Queensland University of Technology (1), OCLC Online Computer

Library Centre (1), University College London (1), Cloud of Data (1) and 2 consultants who are not affiliated to any organisation.

Prior to each interview, and following selection, introductory contacts were made, via email, in order to obtain the consent of each interviewee as well as to reach bilateral consensus as to the timing and venue of the meeting. Due to geographical dispersion and convenience of availability, 14 interviews were conducted remotely (out of which 12 were via Skype and 2 were over the telephone), whilst 7 of the interviews were conducted face-to-face. Once each interview was conducted, it was transcribed and stored in NVivo 9 software (see Figure 7.1).

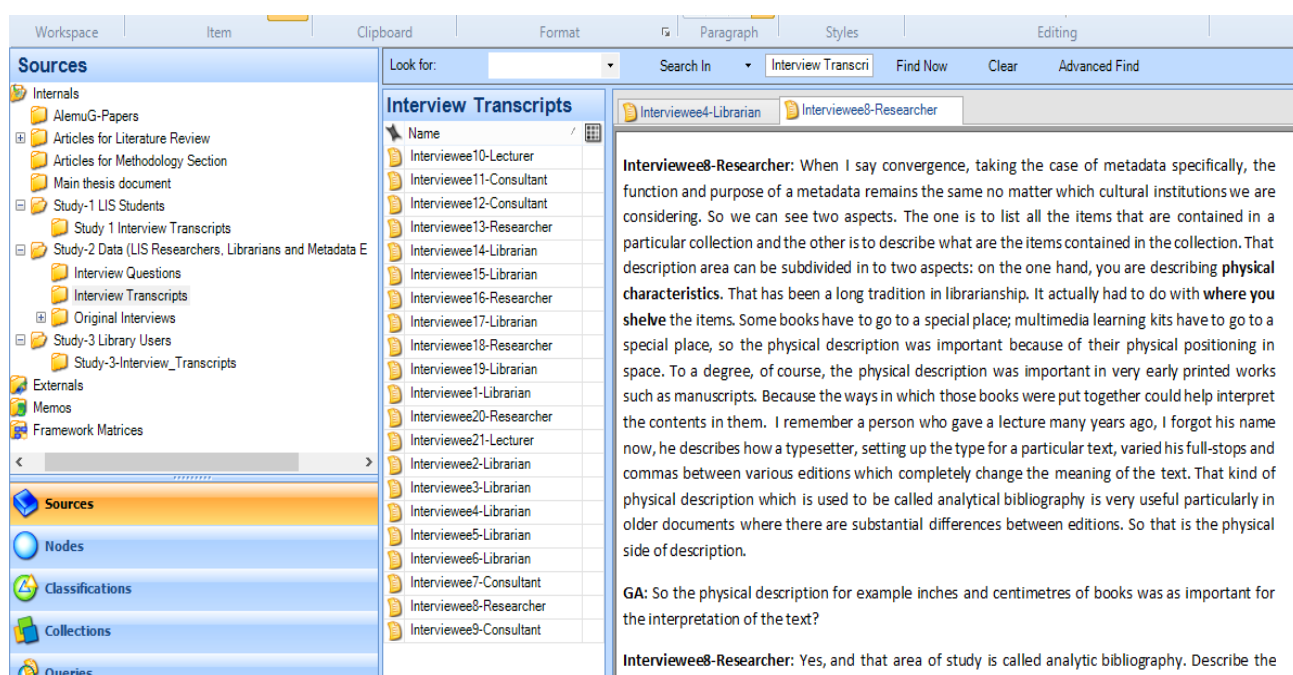


Figure 7.1: Sample Interview Transcript in NVivo 9

After transcription, the interview data was coded, for example Figure 7.2 shows a part of the interview text that discusses the use of the MARC metadata format is coded with an Open Code, i.e. “using MARC”. Any other interview text that deals using the MARC standard, wherever it may appear, is thereafter associated with or given the same code.

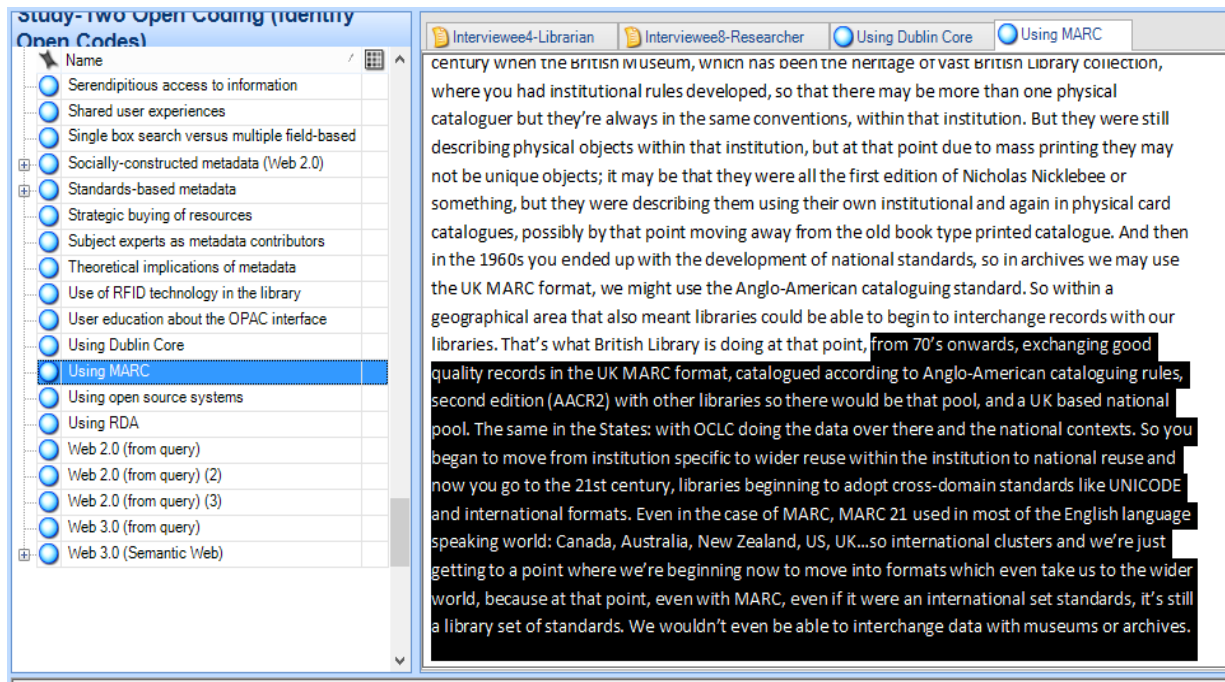


Figure 7.2: Open Coding Stage Using NVivo 9

Using similar procedure as shown in Figure 7.2, the Open Coding analysis of interview data with 21 Library and Information Science (LIS) professionals including researchers, librarians, lecturers and metadata consultants resulted in a total of 221 Open Codes that were deemed pertinent to the representation of the data collected. All the 221 codes that have been established during the Open Coding stage have associated with them a segment of interview transcript, providing a hook to bring together related responses from other interviewees. The establishment of the Open Codes was iterative and evolving. Figure 7.3 shows a sample of these 221 Open Codes.

Open Coding (Identify Concepts)		
Name		
Adding context	Emerging vocabularies	Metadata granularity
Adding structure to social metadata	End of print journals	Metadata interoperability
Allowing users to add metadata	Enriching library metadata with links	Metadata personalisation
Arcane metadata format	Enriching metadata with links and relations	Metadata richness versus simplicity
Associative trails	Ensuring metadata consistency	Metadata sharing
Augmenting legacy metadata with socially-constructed metadata	Ensuring metadata quality	Metadata silos
Author-created metadata	Ethical issues on usage of library data	Minimalist approach to metadata
Blending metadata	Evaluative metadata	Monetizing metadata
Carrying out searches	Evolving metadata scheme	Motivation
Challenges to moving library data to the web-context	Federated search	Motivation for socially-constructed metadata
Changing metadata standards	Filtering using metadata	OPAC
Collections and their metadata as the only constants in libraries	Folksonomies	Open library data
Competing for users' attention	Forsaking the idealised view of the perfect 'user'	Open world assumption
Critical mass	Forseeing future use of metadata	Opening up metadata
Cross-referencing	Full-text searching	Policing
Crowdsourcing	Gaining traction	Polishing metadata
Data-centric metadata	Going to users' spaces	Provenance (metadata on metadata)
Designing serendipity with the digital library	Harnessing computing power to make sense of socially-	Providing alternative spelling suggestions during
Developing new business model using metadata	Having expertise in a specialised domain	Providing better navigation
Disconnect between librarian intentions and user interests	Identifying information use patterns	Providing integrated discovery interface
Displaying metadata	Identifying reasons to adopt new standards	Purchasing metadata records
Document-centric metadata	Idiosyncratic tags	RDA
	Improving discoverability	Representing user vocabularies (terminologies)
	Improving findability	Return investment of metadata

Figure 7.3: Sample Open Codes of the Open Coding Stage

7.2. Use of Terminologies for Coding and Analysis using Sample Codes from the Three Stages of Coding

For the sake of consistency, in agreement with the Constructivist Grounded Theory Method, the following terminologies are used in this thesis.

- **Open Codes:** During the **Open Coding** stage, segments of interview transcripts are coded with labels. The resultant labels are referred to in this thesis as Open Codes (see Figure 7.4). All pertinent responses from all interview data were coded with labels that were considered to be representative.

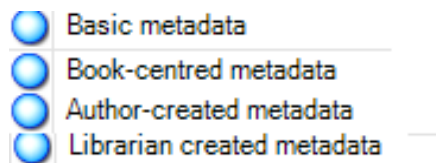


Figure 7.4: Example of use of Open Codes

- **Category:** During the **Focused Coding** stage, Open Codes are categorised into a nested hierarchy. During the early coding (Open Coding) stage, interview transcripts sharing similar themes were labelled with identical codes. Subsequently, closely related codes, in turn gave rise to the identification of broader concepts that more comprehensively represented the underlying issues and phenomenon discussed by interviewees. Finally, the broad concepts were further purged of particulars related to time, place and personality contexts and used to develop higher order hierarchy of constructs, commonly referred to as **Categories** in the Constructivist Grounded Theory Methodology (see example in Figure 7.5).

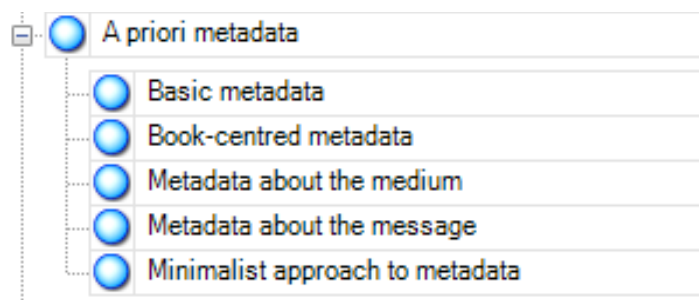


Figure 7.5: Sample Focused Code (Category)

- **Core Categories:** During **Theoretical Coding**, the Categories of Focused Coding are compared, linked and integrated to form what is referred to here as Core Categories (Figure 7.6).

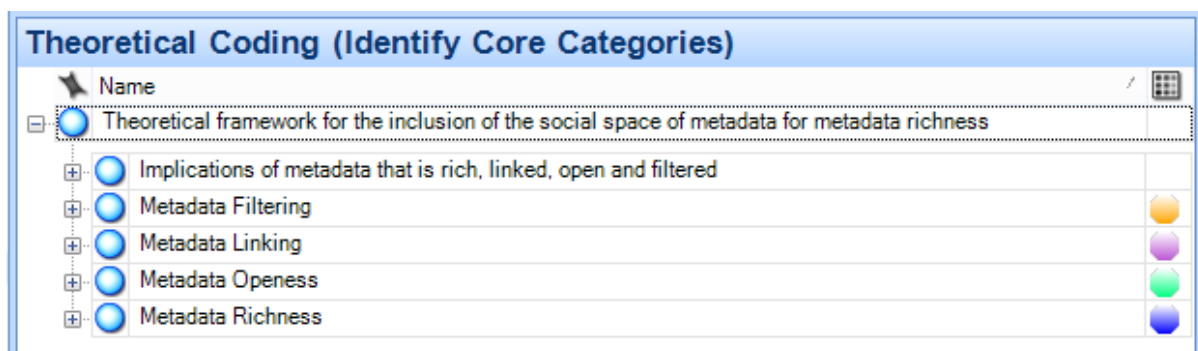


Figure 7.6: Core Categories that Emerged from the Theoretical Coding Stage

7.3. The Current Status of the Adoption of Standards-based Metadata Approaches in Libraries

Broadly speaking, in relation to standards-based metadata approaches, the interview questions for Study Two covered topics on the interviewees' views regarding contemporary metadata standards (such as MARC and Dublin Core), OPAC, metadata quality concerns, as well visions for metadata in both the short- and long-term. Section 7.3.1 provides research narrative of interviewees on the current status of metadata and Section 7.3.2 summarises the perspectives of interviewees on metadata actors. Finally, Section 7.3.3 discusses the identification of 'a priori metadata' as a Category of the Focused Coding interview data analysis.

7.3.1. Research Narrative of Interviewees' on the Current Status of Metadata Standards

Interviewees expressed their familiarity with one or more contemporary library metadata standards. For instance, one interviewee (Study2-Interviewee1) stated that they have been working in libraries since 1973, and recalled their experience in utilising pioneering subject headings and classification systems, such as the United Kingdom's MEDLARS (Medical Literature Analysis and Retrieval System), the National Library of Medicine's Medical Subject Headings (MeSH), MARC and the Dewey Decimal Classification System. Another interviewee (Study2-Interviewee6) stated their experience working with DDC and MARC as far back as the early 1980s. Similarly, a third interviewee (Study2-Interviewee3) alluded to the fact that they are currently using MARC as a metadata creation format and the Anglo-American Cataloguing Rules second edition (AACR2) as their guideline for describing and cataloguing information objects in their library. In addition, interviewees mentioned their familiarity with Dublin Core, which was depicted as a simple descriptive metadata standard.

Furthermore, most interviewees disclosed being aware of what some of them termed "emerging" and "novel" library standards and principles, such as FRBR and RDA. Some interviewees, however, believed that there still existed gray areas as to whether these new approaches were capable of accommodating the evolving and changing needs of users. They also believed that these approaches were described and modelled with both verbose and

lengthy documents (mainly RDA) and were characterised by the use of jargon and complex language (FRBR). One metadata consultant (Study2-Interviewee11) critiqued RDA by saying, “its actual approach to data is very similar [with previous cataloguing rules], i.e. you have one record per book, it has a very similar structure even though it looks different, and a typical library, this is a cynical remark, would spend ten years building the RDA standard, and the major deliverable is the RDA manual”. Yet, another interviewee (Study2-Interviewee9) also added that “RDA is still using text-based entry points”.

In connection with this, two interviewees who are metadata consultants (Study2-Interviewee7 & Study2-Interviewee11) voiced their concerns with regard to the capabilities of FRBR and RDA in addressing current metadata challenges. One of the interviewees (Study2-Interviewee7) jokingly noted that looking at FRBR “is not fun”, alluding to the various FRBR component entities, such as Work, Expression, Manifestation and Item (WEMI) and the complex relationships between these components. On the other hand, one interviewee (Study2-Interviewee19), who once served as the chair of the Joint Steering Committee for the Development of RDA (JSC), and is currently coordinating its implementation at their library, noted that “RDA was only intended for consumption by professional cataloguers so you wouldn’t expect anyone who is not a librarian to be at ease in using RDA”.

Yet, another metadata consultant (Study2-Interviewee9), who participated in the initial development of RDA, admitted the fact that many people have found both RDA and FRBR complicated. The solution, as the interviewee added, is to jump altogether to a Linked Data model and then map the resultant metadata from Linked Data to FRBR, thus addressing the objectives of FRBR to support users to find, identify, select and obtain information objects. This was not a viable approach given the current attempts which are geared towards Linked Data. One interviewee (Study2-Interviewee15), whose institution is involved in a major initiative to develop a new bibliographic framework, alluded to the importance of moving from a document-centric metadata environment to granular and machine process-able metadata description, which focuses on relationships. Only then, the interviewee believed, can libraries fully implement the relationships between the various entities in the FRBR model. It is important to note that, whilst FRBR was designed as a conceptual bibliographic mode, RDA (replacement for AACR2) was aimed at instituting cataloguing rules which could help guide the implementation of FRBR. One metadata consultant (Study2-Interviewee11) observed, “when they [library system developers] want to build systems using the FRBR model, they discover the data in the MARC record was not good enough to populate a FRBR model without too many anomalies”, thus the RDA initiative came into being.

It is also noteworthy that, IFLA has developed the FRBR model with the aim of enhancing the findability, identify-ability and accessibility of information objects, which inherently have complex structures. The complexity emanates from an information object’s being an intellectual work possessing multiple instantiations, through being published in various mediums, editions, and being an identifiable item on a particular shelf or network location, thus necessitating the componentisation of an information object into WEMI (see IFLA, 2009 for details on the model).

Unfortunately, as the interviewees indicated, not everyone is happy with the FRBR structure. Thus, one interviewee who is a metadata consultant had the following to say: “The real world isn’t as formally structured as librarians would like. No matter how well structured FRBR is, people don’t know whether something is Manifestation or a Work or Item” (Study2-Interviewee11). The issue of jargonising has also been raised by other interviewees, wherein they alluded to the importance of simplifying these models so that, not only librarians, but external agencies and users can also become well aware of them and use them appropriately. As one interviewee (Study2-Interviewee21) indicated, both FRBR and RDA, which are nonetheless considered new or emerging metadata models by many in the LIS domain, are becoming old and obsolete. In their own words, the interviewee opined thus, “libraries have come a long way, from rule oriented AACR to RDA and now Linked Data (the BIBFRAME model). But even now these efforts which were considered as emerging trends are considered as past efforts as we move in to a Linked Data environment”. Whilst still maintaining their argument about the relevancy and novelty of FRBR and RDA, one interviewee admitted that, the area of library standards is in a state of continuous change. “I think this is quite an exciting period. Everything is in flux. It is a moment of great opportunity for libraries but also a significant risk if we miss this opportunity. The opportunity may not come to us again” (Study2-Interviewee19). This view was supported by another interviewee (Study2-Interviewee7) when they stated their strong belief that “the library world is considering undertaking a large change”. The interviewee indicated their wish to see this change result in thinking beyond the library walls. One of these, as the interviewee believed, is to be able to create re-usable library metadata that can be used within and outside the context of the online catalogue.

As the interviewees showed, contemporary metadata standards are thus well established in libraries and the people working in them, mainly librarians, base their operations using these authoritative approaches. This was evidenced in interviewees’ discussions by naming the metadata standards during the interview process, which showed their familiarity and awareness.

As to the job titles of some interviewees, these included: metadata expert, head of metadata services, metadata consultant, metadata researcher and lecturer on standards. For instance, one interviewee (Study2-Interviewee3) had the following to say about their job description: “formerly employed as a bibliographic resource officer, I am now what they call a metadata librarian. I do mostly traditional cataloguing using AACR2 (Anglo American Cataloguing Rules, Edition 2) which uses the MARC coding format. I have also worked with other library staff on metadata schemas, mainly Dublin Core, which is used in the e-prints repository”. The interviewee further added that their work also included ensuring whether metadata that comes from external sources matches the right format, checking for missing metadata, and ensuring consistency and accuracy. Another interviewee (Study2-Interviewee4), who is head of metadata services at a national library and has been involved in several international metadata projects, including participation in the development and improvement of standards, described at length the progress from card catalogues, to electronic metadata formats using MARC, to the coming into existence of simple descriptive standards such as DC and ONIX (metadata

standard used by book publishers), authority lists (MARC country and language lists) and subject headings (MeSH and LCSH).

Whilst the interviewees tended to point out that MARC is antiquated, especially in the light of changing users' needs and external impetuses, such as data sharing and re-use, nonetheless, they all concurred that there is no easy and fast way out of MARC. Some interviewees were aware that the Library of Congress, the originator and maintenance agency for MARC, is itself coordinating efforts geared at replacing it with a web-compatible and link-friendly metadata format. Many interviewees pointed out that the MARC standard has served its time, indicating thus the importance of changing it. Whilst many of the interviewees pointed to the limitations of MARC, another interviewee, whose institution is responsible for the development and maintenance of MARC since the 1960s, recognised that:

“I think in terms of moving from MARC to new web-based models there has been sufficient research pointing to the anachronistic of MARC. Because we have been told since the last ten years [since 2002] that MARC needed to be replaced. But the move towards RDA brought us to a concrete need for change, because we could not implement RDA with MARC. We got actual feedback from cataloguers who do the job. Because they said yes we could implement RDA but we will not have much benefit since the end result looks the same. Cataloguers believe RDA can be optimally implemented with new metadata structures that are based on relationships. We received feedback that MARC is not enough and that a robust metadata model is required particularly for the digital and web environment” (Study2-Interviewee15).

Thus, the interviewee believed that, it is only time that MARC, which has persisted since 1960s, needed to change. One interviewee (Study2-Interviewee10) pointed out his belief that a “successor standard [to MARC] is going to arise”. The interviewee added that, “so much has happened in those years. So it is time that we design a new metadata structure and while we do that we need to deal with the millions if not billions of legacy MARC data and shift to the new models and formats”. Most interviewees also noted that they had always thought MARC could have disappeared long ago, given its age and the changes in technologies and formats. It seems safe to assert that MARC prevailed partly due to the absence of new standards that could carry existing records forward through a smooth transition. The interviewee, nonetheless, remarked that MARC had some notable limitations. Another interviewee revealed that their national library still uses MARC because they found it versatile. In addition, “we have got a lot of tools [utilities] that enabled us to do things like sorting and filtering and they all use MARC” (Study2-Interviewee4).

The existence of tools and software that better handle and extract data from MARC records strengthens the argument that, unless tools and applications are developed for making sense of the data and manipulating it in the upcoming alternative formats, there would be sluggishness, even reluctance, by libraries to introduce these new formats. The interviewee remarked that contemporary library software and applications have been designed to retrofit MARC. The same interviewee added that as far as they were concerned MARC offers “a better degree of richness and sophistication and detail” than other formats, such as DC or

MODS, but acknowledged the limitations of MARC, particularly in relation to facilitating library metadata accessibility and usability by applications (machine process-ability) and share-ability with other libraries.

On the other hand, yet another interviewee, a metadata consultant, maintained that MARC had to change, stating:

“I think they will have to change. The work that IFLA is doing to take the standards formats such as ISBD, UNIMARC and put it in RDF is not going to get us where we need to be. It is an interesting exercise but it isn’t really producing [the desired result]. It is much like what we did when we moved from the card to the MARC, wherein we carried over the same data in the same way into just a new format. What we need to do is to change the way of thinking about our data. So I kind of understand IFLA doing this but doing just that I don’t think that moves us very far forward” (Study2-Interviewee7).

It is important to note that at present the Library of Congress, the British Library, IFLA and many other national libraries and international agencies are discussing the question of how to move to newer standards that can replace MARC. The Library of Congress, in particular, is taking the initiative for developing a new bibliographic standard.

All interviewees singled out MARC as being well established throughout the world, as it, having undergone several iterations, has a number of instantiations in several languages. In fact, they all seemed to agree that, library metadata could be considered synonymous with MARC, even though the latter only refers to a single standardised format. Most library software have been designed in such a way that they accommodate and/or are cognisant of the MARC structure.

All the interviewees professed that, since MARC is the dominant descriptive metadata standard, there are hundreds of millions of records (one conjecturing more than 800+ million) in libraries throughout the world. One interviewee (Study2-Interviewee10) was re-assuring when they stated that a significant amount of library bibliographic data created using the MARC structure will not simply vanish or remain unusable, indicating the possibilities of mapping it into various other formats. The interviewee emphasised the conceptual underpinnings of any standard including integration, interoperability and granularity. The interviewee for instance opined that conceptual themes such as semantic and structural relationships should be a main point of concern for discussion as opposed to technical instantiations of these concepts. For this interviewee XML and RDF are all technical formats, thus suggesting that some of the discussions about the power of RDF is overrated. “So [RDF] is not a matter of revolution, it is a matter of alternative formats” (Study2-Interviewee10). Citing a recent project they have undertaken, the interviewee added that schemas previously designed using XML can easily be mapped into RDF and vice versa.

In contrast, the responses from four metadata consultants (Study2-Interviewee7; Study2-Interviewee9; Study2-Interviewee11; and Study2-Interviewee12) seem to position RDF as a higher level data modelling language than a mere technical format. However, the data

analysis revealed a subtle, but nonetheless important difference between those that support the move to RDF. For example, one of them passionately considered RDF (Study2-Interviewee11) as a panacea to the issue of document-centric metadata and criticised libraries as slow to adapt, another (Study2-Interviewee12) whilst still supporting the use of RDF warned that “Yes, there is a place for it. It does add some value. I think it would be dangerous to get dogmatic about it and to start making everything RDF”. The first interviewee thought the cross-mapping shift from MARC records, to MARC/XML and recently from MARC/XML to MARC/RDF was piecemeal and, as the interviewee added, even though some librarians thought that is a significant change, it failed in practise to change the document-centric metadata view to one that is data centric and machine process-able, thus maintaining the status quo of contemporary standards. The interviewee believed the format changes as putting the old data into new data covers (envelopes), the content remains the same. But another interviewee (Study2-Interviewee9) thought the pragmatic translation between these various standards would slowly help to move the library domain from its silos into web compatible approaches. Because, as the interviewee believed, maintaining the status-quo “is going to destroy the library profession. Because other RDF-based, communities including users are going create their metadata in triples, social networking sites already allow users to create their own metadata and trending towards RDF”.

Looking at alternative formats, most interviewees estimated that it would take a number of years until contemporary legacy library metadata is fully migrated into newly developed ones. One interviewee (Study2-Interviewee15) put their estimate from two to five years. Most interviewees, nonetheless, agree that there would be a parallel use of the old and new formats until such time that all the records from the old are fully transferred into new ones. One interviewee even added the possibility that some libraries may even prefer to continue to use MARC as it is “fairly sufficient for what they do” (Study2-Interviewee15). Whilst the views of interviewees on the speed and extent of changes differ, all of them agreed to the enormity of contemporary bibliographic records embedded through old formats and the resource, skills and technological implications of data migration and cross-mapping.

The interviewees remarked that, in order to smoothly change contemporary metadata standards, the issues of legacy metadata should be clearly and satisfactorily settled. There are, as the interviewees underscored, significant amount of records using contemporary formats. In addition, library systems and applications should equally be re-scrutinised. In general, the interviewees opined that metadata standards have an important role in the metadata creation and utilisation functions of libraries. “In an ideal world, as a librarian, I would like to see everything standardised. I think there are some places where it becomes awkward if you don’t standardise. For example, if you don’t have standards for author names, it can become frustrating” (Study2-Interviewee6). There is consensus on the part of the interviewees, especially among librarians, that standardisation facilitated international and regional collaboration on metadata record creation and sharing. One interviewee, in particular, emphasised the importance of standardisation for metadata interoperability (Study2-Interviewee10). This, however, implies that libraries have to abandon their localised formats, having implications on accommodating local metadata needs. Libraries do not seem to be in a

position to seamlessly integrate local and international schemas. Thus, as one interviewee (Study2-Interviewee6) noted, their library was forced to scrap the local ones in favour of international standards. In other words, standardisation, in most cases, obviates the use of local metadata schemas, in favour of internationally (or nationally) approved ones.

As the narratives from the interviewees above indicated, the current library metadata functions are characterised by the use of standardised sets of metadata elements, controlled vocabularies and encoding rules that are purposefully designed to ensure consistency and accuracy of metadata entries, which would in turn ensure the quality of metadata (detailed results are shown in Section 7.4). The emphasis on standardisation has implications on metadata creation and utilisation. One such implication is who should be creating metadata?

7.3.2. Perspectives of Interviewees on Metadata Creators in Current Metadata Practises

According to the interviewees' responses, whilst there are four types of metadata actors identified, which include metadata experts (librarians), authors, machines and users (see Figure 7.7); among which, metadata experts are the predominant, if not the only, ones who assume the responsibility of creating and managing metadata at present in most libraries. Interviewees emphasised the lead role of librarians, as metadata experts, in creating metadata records anew, curating metadata records acquired from commercial or third party vendors, and checking author-created metadata for accuracy. Author-created metadata is only minimally used in libraries, although it is beginning to gain some acceptance in the context of institutional repositories. One interviewee noted the challenges of author-created metadata, especially in the context of obtaining granular subject-headings for published articles. Thus, librarians are heavily involved in standardising the metadata assigned by authors, along with their to-be-published articles. Consequently, metadata experts (including experts from external service providers) are the major actors in the metadata creation and maintenance functions. The interview data revealed the minimal significance of machine generated metadata as well as the total absence of user-created metadata in widespread usage. Although the potential role of socially-constructed, i.e. user-created, metadata approaches will be dealt in great detail in Chapter Eight (8.1), machine generated metadata is out of the scope of this research.

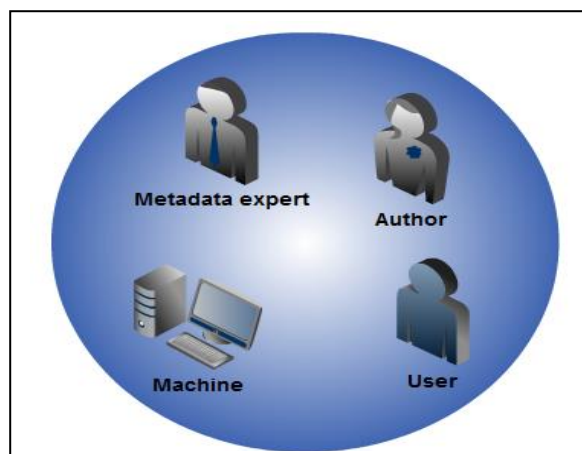


Figure 7.7: Potential Metadata Actors

The metadata created by metadata experts (librarians) mainly fits a predetermined schema of the standards adopted. Some interviewees call this basic metadata. Being basic refers to its importance in helping users to find and retrieve information objects. Figure 7.8 shows the dominant type of metadata relationship currently in use in libraries.



Figure 7.8: Metadata Experts as Major Actors Creating/Curating Basic Metadata

The Open Coding stage resulted in the identification of a list of codes pertaining to the creation of basic metadata. These codes were then collated together into a hierarchy. These open codes included: book-centred metadata, metadata about the medium, metadata about the message, and minimalist approach to metadata, which thus were categorised as basic metadata. As one interviewee (Study2-Interviewee8) pointed out, for a long time the function of metadata has been considered from the point of view of describing the physical carrier of information objects. This is a characteristic facet of “old fashioned” cataloguing, wherein the main operation was the description of the medium, but not the information or actual ideas contained. The interviewee was of the opinion that librarian-created metadata is thus mainly focused on describing the characteristics of the medium (carrier), and not the information (ideas) contained in the medium. The interviewee noted the merits of attaching such metadata, especially when an understanding of the medium adds value to the better understanding of the content within it. The interviewee further stated that describing the content requires additional knowledge of the subject matter, thus suggesting a role for other metadata actors, such as users.

It is important to note that the interview text with regard to the above codes was mainly in reference to librarian-created metadata. Basic metadata is considered essential, in order to make information objects findable and discoverable by users. Some interviewees, however, refer to basic metadata as the process of description of information objects with simple and minimal metadata, which they think is made necessary as the size of information objects, especially digital ones, continues to grow. For instance, one interviewee (Study2-Interviewee1) commented “I am a bit of a minimalist, I would prefer we invest time on better metadata in many ways”. However, many other interviewees argued for detailed metadata, provided that the current challenges are surmounted with alternative approaches.

As the interviewee quoted earlier indicated, basic metadata can be instantiated or created using standards-based metadata approaches. Another interviewee (Study2-Interviewee3) described basic metadata as “metadata created beforehand, metadata which librarians create

as surrogates of the resources, so as to enable users to find resources and determine whether it is useful. As pointed out by Study2-Interviewee3, the synonym for basic metadata is core metadata, indicating its unique importance as a provider of the initial hook to information objects, meant to be found and discovered by users.

7.3.3. A Priori Metadata

Using the Constructivist Grounded Theory coding and using the NVivo 9 software, interview excerpts/quotes (that are related to the experiences of LIS professionals in their use of standards were initially labelled with Open Codes such as **“using MARC”**, **“using Dublin Core”**, **“using RDA”**, or **“using Subject Headings”**. The interview data thus labelled with these codes describes the current use and adoption of metadata standards in libraries. Other codes that dealt with strongly related themes were brought closer to these Open Codes. Those codes that were identified as having a similar theme as that of **“The Current Status of the Use of Metadata Standards”** were: **“Librarians as Metadata Creators”**; **“Authors as Metadata Creators”**; and **“Basic Metadata”**. It is worth noting that the interview quotes associated with these codes were found to exhibit an implicit relationship among them. This is because the interview excerpts clearly allude to the fact that librarians are the main metadata creators, followed by authors who attach metadata to their works in the context of institutional repositories. As shown in Figure 7.9, the direction of arrows demonstrates how interview quotes were coded during the Open Coding stages, which then were used to develop more focused constructs, called Categories, during Focused Coding.

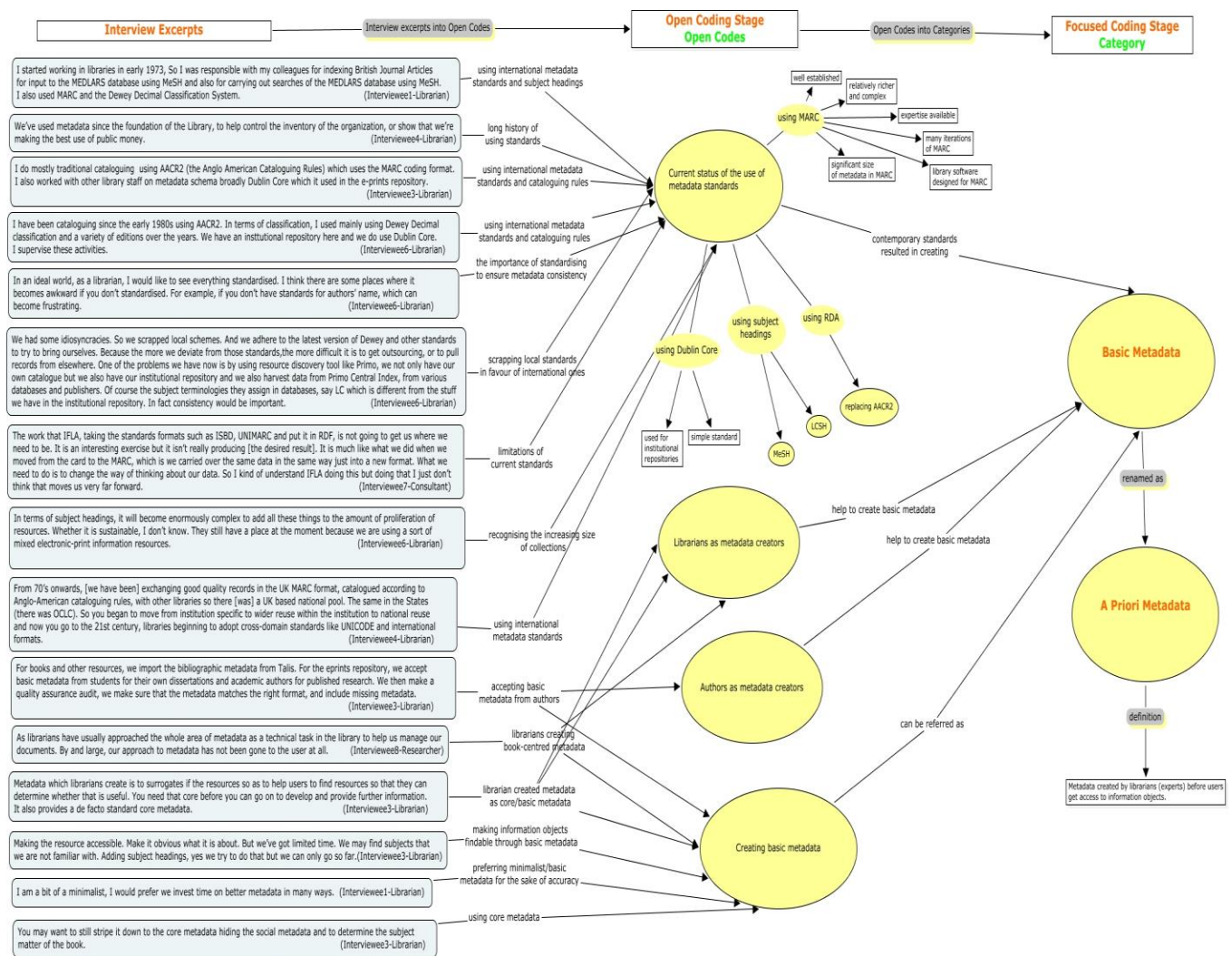


Figure 7.9: Data Analysis Showing the Emergence of the Concept of “A Priori Metadata”

Following a closer examination into the nature of basic metadata and its value as a core set of metadata, in this thesis, the term “a priori metadata” is considered a more explanatory nomenclature. This is also in accord with the Constructivist Grounded Theory Methodology where terms that can have grab and resonance can be preferred. Figure 7.10 illustrates a conceptual relationship between standards, a priori metadata, information objects and metadata experts.

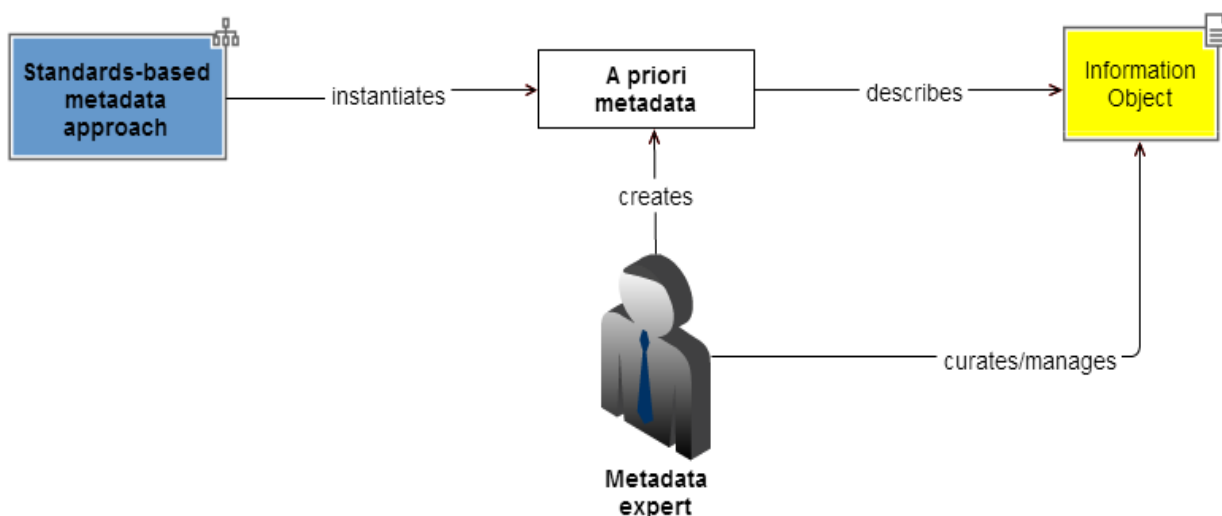


Figure 7.10: Standards-based Metadata Approaches Instantiating A Priori Metadata

When it comes to the process of metadata creation, it is important to bear in mind that present-day metadata standards implicitly presume that authors create works, that librarians create metadata, and that users access information objects. Hence, standards-based metadata is predominately generated a priori, i.e. before users get access to a particular information object, such as a book. In other words, the sequence of steps is as follows: librarians acquire an information object, describe it with metadata based on a given schema (or acquire metadata from third party suppliers, such as Online Computer Library Centre, Inc. (OCLC), who, incidentally, prescribe to the same rules and regulations), and finally make the information object available on library shelves or in electronic databases so that users are able to access and utilise it. It is in rare cases, such as in self-archiving institutional repositories, that a priori (before-the-fact) metadata is created by authors of the works themselves.

A priori metadata can be defined as the metadata created and affixed to an information object before “publication”, i.e. before release of information to users. It can thus be considered as before-the-fact metadata, as the description of the information object is made before its use. It thus has implications on metadata creation as poor a priori metadata or the absence of a priori metadata would negatively affect the findability of an information object.

7.4. Open and Focused Coding Relating to Standards-based Metadata Approaches

In addition to examining the current status of standards as discussed in the research narrative in sections 7.3.1 and 7.3.2, further analysis of the interview data using **Focused Coding** also revealed a number of **Categories**, which need to be considered when developing future metadata standards. The categories that emerged during the said stage include: metadata structure (7.4.1), granularity (7.4.2), provenance (7.4.3), quality (7.4.4), simplicity (7.4.5) and interoperability (7.4.6). Each of these **Categories** is discussed below, along with the examples of how each was derived from the interview data.

7.4.1. Providing Metadata Structure

Interviewees discussed standards in terms of their constituent elements, the use of controlled vocabularies and the associated metadata (cataloguing) rules that are enforced when creating records. When describing information objects using such standards, metadata experts either create the data values by referring to the information object itself (for example Author, Title and Year of Publication can be recorded from the title page of a book) or such values can come from an existing list of controlled vocabularies. Furthermore, the recording of these values adheres to an established encoding rule or constraint so that the values are consistently entered in a particular pre-defined format. Considering the interview data, during the Open Coding stage, resulted in the identification of three major codes ‘Controlled Vocabularies’, ‘Encoding Rules’ and ‘Metadata Elements’ (see Figure 7.11).

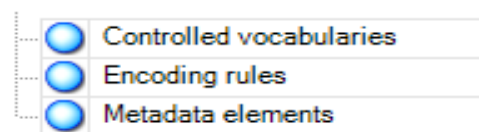


Figure 7.11: Open Codes relating to Metadata Structure

Since the three **Open Codes** deal about one common theme, they were subsequently grouped under the **Category** called ‘Metadata Structure’. To make the category self-explanatory, in accordance with the Constructivist Grounded Theory Methodology, an active verb (gerund) was added, resulting in the category, ‘Providing Metadata Structure’ (see Figure 7.12).

The presence of these three components enables a metadata standard to provide the basic structure required for describing information objects. One interviewee (Study2-Interviewee3) identified the main concerns of librarians being that of metadata consistency and accuracy. The interviewees tended to agree that standards-based metadata approaches are the primary approaches for organising information objects within a given semantic and syntactic structure. Some interviewees, all of them affiliated to libraries, indicated that standards-based approaches have an important role to play, especially in providing structure. These interviewees singled out in particular, the role of controlled vocabularies, such as subject headings, as important components to ensure the structure of metadata. However, it should be borne in mind that the role of subject headings is particularly appropriate in specialised domains, rather than in broader and multi-disciplinary ones. For example, with regard to subject headings, one interviewee (Study2-Interviewee1), described their experience in the highly specialised area of medicine, whereby Medical Subject Headings (MeSH) are used to add metadata values from a list of controlled subject headings. Such metadata, the interviewee noted, that controls subordinate metadata values tends to be more applicable in specialised and well developed subjects such as medicine. Hence, MeSH has evolved to become a full-fledged controlled vocabulary in the field of bio-medicine. In contrast, the development of subject headings in multi-disciplinary topics tends to be more complex.

As the interviews revealed, one of the most notable strengths of standards-based metadata approaches is its ability to provide an underlying structure. For example, classification schemes and thesauri bring related subjects together through hierarchical and associative

relationships. Metadata schemas provide a predetermined set of metadata elements, so that values are added to the appropriate elements, thereby providing both structure and consistency. In addition, the use of controlled vocabularies and authority lists, whereby values are selected from a predetermined list, is considered an important strength. The data analysis indicated that standards-based metadata are defined chiefly through the existence of metadata structure that has three interrelated components, namely metadata elements (fields), controlled vocabularies and encoding rules. A library's decision when adopting a particular metadata standard is highly dependent on, among other things, these three components of the standard. The number of elements is an indicator for checking the completeness of a metadata description. Overall, metadata structure is an indispensable competency, helping to ensure the consistency and accuracy of metadata records. However, an optimal balance should be struck with regard to the range of metadata elements, selection of controlled vocabularies and choice of encoding rules so that the structure does not inhibit the description of information objects. One interviewee, for example, mentioned that some metadata standards are inadequate for describing information objects due to their limited constitutive element structure, citing the case of MARC (with more than one hundred metadata elements) and Dublin Core (with only fifteen).

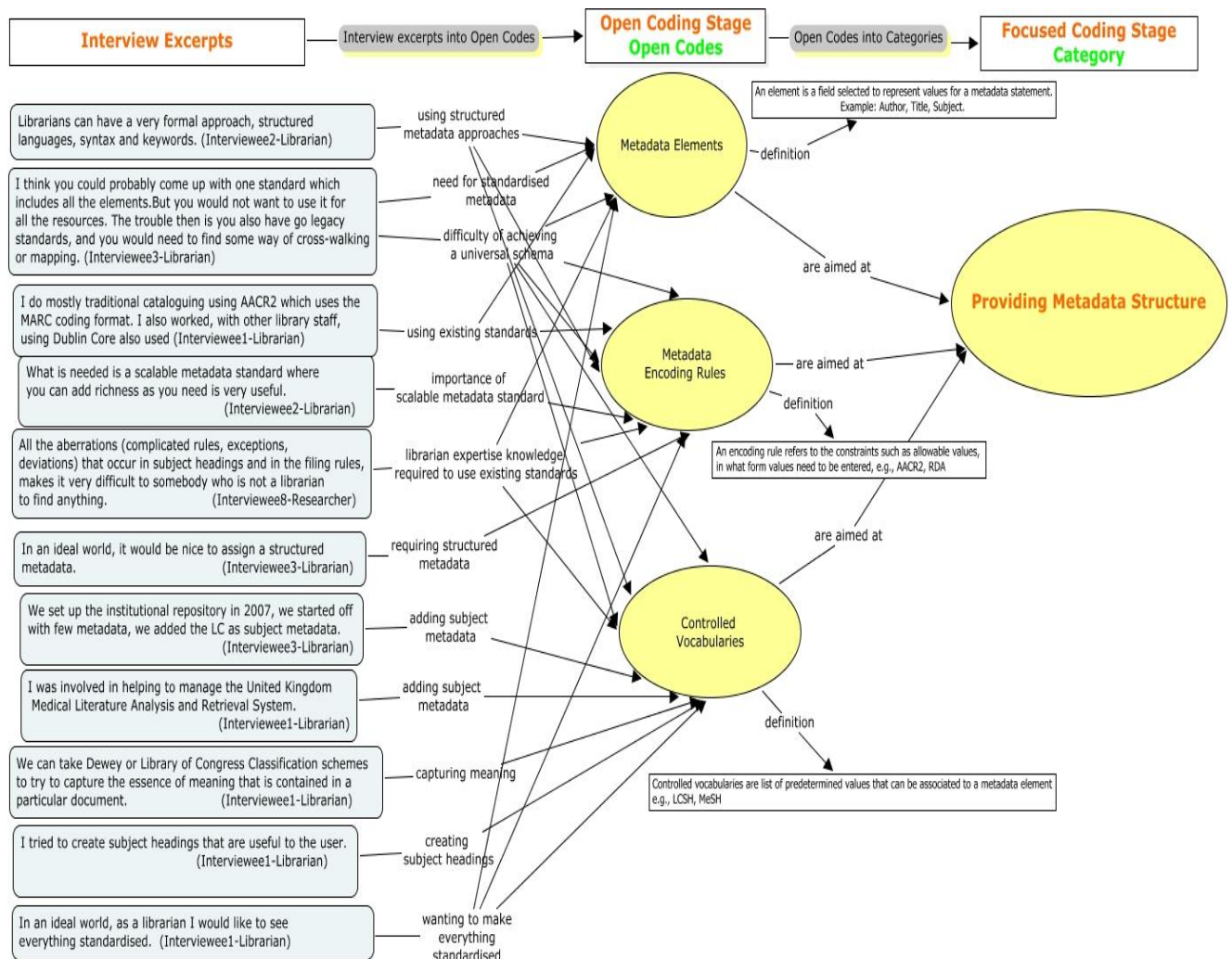
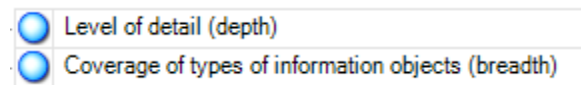


Figure 7.12: The Category of “Providing Metadata Structure” during Focused Coding

7.4.2. Providing Metadata Granularity

Interviewees considered granularity in two major dimensions: depth and breadth (see Figure 7.13). These two dimensions were identified during the Open Coding Stage of the analysis.



●	Level of detail (depth)
●	Coverage of types of information objects (breadth)

Figure 7.13: Open Codes relating to Metadata Granularity

Depth refers to the level of specificity and detail when describing an information object. In connection with this, one interviewee (Study2-Interviewee1) expressed their concern over what they called broad metadata that lacks detail. In order to include depth of coverage, the interviewee believed, greater efforts should be expended when describing information objects and/or their constituent parts. The interviewee further pointed out that some domains, such as biology, have very highly specialised areas and thus an information object, such as a book, having in its subject field the term ‘Biology’ provides little information. The interviewee, while acknowledging the current challenges to having such rich descriptions, noted that users are more and more interested in such depth.

Another interviewed noted:

“If you are now having to work at the chapter or the article level and create effectively a miniature piece of descriptive metadata for that work in the way that you once had to do for the entire book or for the entire journal, [you would now need to go in much finer detail to describe each of the component parts]” (Study2-Interviewee4).

The interviewee, however, noted the possibility that some users may still opt for metadata that is broad and relatively “superficial”, thus indicating the importance of taking into consideration the requirements of different user categories. It was also noted that a user’s needs may also be dependent on the context of use of information objects. In connection with this, one interviewee noted differences attributable to the typology of a library and its principal users, citing such examples as university and research libraries where greater level of granular description may be required as opposed to public ones.

The second dimension of granularity is the spectrum of coverage or breadth of information. Breadth indicates the range and extent of metadata, including the coverage of multiple types/genres of information objects. Breadth is especially important in light of the complexity of digital objects, which have various components in terms of file types and formats. Metadata should show the links and connections between the various parts of an information object. One interviewee cited the case of e-books that may consist of hundreds of text pages, images and video files, thus suggesting the importance of ensuring such complex objects are described in a manner that supports their findability (Study2-Interviewee4). Interviewees noted that standards-based metadata approaches, in principle, provide better granularity; however, they also pointed out the practical limitations of current standards.

In relation to current standards-based metadata approaches, as two interviewees noted, Metadata Encoding Transmission Standard (METS) provides better granularity as its structure enables the mixing and matching of various metadata elements from disparate schemas, thus providing a level of detail that may be lacking in any one particular standard. The interviewees also singled out the flexible XML structure used by METS as an added benefit. However, it was indicated that this is a library-centric standard and, hence, was not in widespread use by third party information providers. On the other hand, the MARC standard, although having more than a hundred metadata elements, lacked a link structure, while its tags, as many interviewees noted, were listed in a flat file format, and was thus considered limited in its granularity. Web-based standards, such as Dublin Core, are designed with a very simple structure, and thus lack depth. In contrast, as one interviewee (Study2-Interviewee7) remarked, Linked Data models provide atomic granularity as they allow the inclusion of metadata at any level, without being constrained by the number of fields, unlike current standards (see 10.4).

In connection with subject headings, one interviewee remarked that existing subject headings were too generic, indicating the importance of more granular descriptions of concepts. It was indicated, as the collections of information objects increases, broader subject term searches are likely to result in high but irrelevant recall rates. According to one interviewee, what is required in libraries today is a standard that provides a greater degree of granularity that goes deeper into the structure of the document, down to the article and chapter level. Whilst some publishers are attempting to provide full-text chapters in electronic formats, such as e-books, granular metadata is likely to remain essential. However, as one interviewee (Study2-Interviewee4) pointed out, libraries are not yet ready to offer this kind of metadata. The main challenge is the enormity of the task, when attempting to annotate individual chapters with pertinent metadata. This is perhaps where libraries may want to involve users (8.3).

It is important to note that adding more metadata granularity has resource implications. The interviewees indicated that adding even a single metadata field requires huge investments, as the new structure affects millions of information objects. Similarly, changing standards also has parallel implications. One of the questions that interviewees wondered about was who would be creating all this granular metadata and who would bear the costs.

Granularity could be looked at from the point of the structural components of an information object; but could also be approached from the point of the number of available metadata fields in a metadata schema. The plethora of current metadata standards is a consequence of the diversity of metadata needs that are required by various domains. Some metadata standards have very few fields, each capable of holding a corresponding single value, whilst others contain a large set of metadata fields, again each field accommodating a single value, while still yet others cover only a few information resources - all indicating perhaps a lack of breadth. Some standards are able to cover only a limited genre of information, such as Visual Resources Association's (VRA) Core Categories for describing visual works of art; IEEE's Learning Object Metadata (LOM) for describing educational resources and Encoded Archival Description (EAD) for archival materials. MARC and MODS are also used for mainly

describing books. Dublin Core (DC) is used for describing general electronic resources. Each metadata standard provides different levels of coverage of the type and genre of resources.

The importance of granular metadata description emerged, from the analysis of data, as an important category of standards-based metadata approaches. Interviewees pointed out that a greater degree of granularity and detail implies greater effort in both metadata design and implementation. It is important to note that, at a more technical level, the METS metadata structure addresses part of the problem of granularity. However, it is also very essential that granularity is looked at a more conceptual and theoretical level. The category emerged from the Focused Coding stage of analysis (see Figure 7.14).

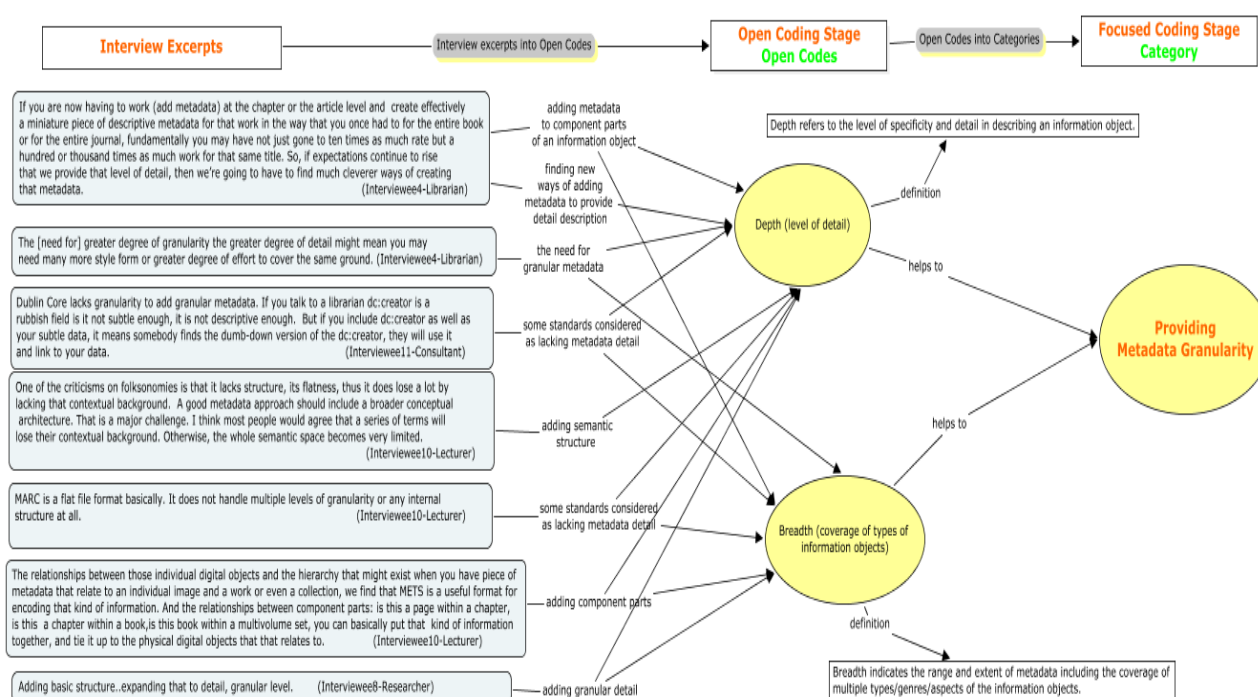


Figure 7.14: The Category of Metadata Granularity

7.4.3. Providing Metadata Provenance

Interviewees considered provenance as an important area of concern for metadata functions. One interviewee (Study2-Interviewee7) stated their belief that provenance metadata informs users about the sources and origins of metadata, thereby contributing to ensure its credibility and veracity. The interviewee noted that through provenance, the user would be able to determine where a specific information object has come from. One interviewee (Study2-Interviewee8) contextualised the discussion of provenance with Paul Otlet's idea of the social space of documents (see Wright, 2007) and the notion of describing the essence of a document. The interviewee noted thus "The context in which the document was created is, and has long been, considered a very important principle in archival work, the whole issue of provenance is related to that. Where the document occurs in a sequence, which organisation it comes from, helps interpret its meaning". The interviewee indicated that the context in which

the metadata was created has implications on its future use. In this regard, the interviewee noted that the institutional (organisational) context of metadata (for example is the metadata created and used in a gallery, museum, library or archive) has significant effects in its use. In connection with this, the interviewee offered the following example:

“I remember I went with a friend for the very first time to the National Gallery in London; my friend went to view a piece of artefact in the Gallery and said ‘I wouldn’t hang this or that in my toilet’; but once she understood from where that piece of art originated as well as some of the technical aspects of its provenance, her attitude towards the artefacts we were looking at became quite different” (Study2-Interviewee8).

Other interviewees also identified source and context of metadata as important constituents of ensuring the authority and authenticity of metadata. From the Open coding stage the following Open Codes were identified (Figure 7.15):



Figure 7.15: Open Codes relating to Metadata Provenance

Thus, as emerged from the Focused Coding stage of data analysis, metadata provenance includes authority, ownership and related contextual information in relation to the metadata that describes an information object. The interviewees indicated that metadata provenance helps in determining the authoritativeness of the source of data, so that users are able to know the reliability of the metadata. One interviewee (Study2-Interviewee9), stated that: “You have to know who said this, when they said it, why they said it, what rules they were applying, what system they used to create the metadata, because systems are going to change”. However, the same interviewee also noted possible variations in the need for the amount of provenance information. Whilst some users might be simply interested in the data, others may need only the source of the metadata, whilst still others would want to verify detailed provenance information. Thus, the interviewee suggested that the presentation of the provenance metadata should be catered to according to a user’s needs.

By bringing the Open Codes of metadata authority, context, history and ownership, the Focused Coding analysis showed that recoding the provenance of metadata is crucial, thus the identification of the emergence of the Category of Metadata Provenance is an important exercise. One interviewee’s (Study2-Interviewee8) remarks summed up the importance of provenance metadata, “no document exists by itself, not only is it linked to other documents, but to other ideas. There are all kinds of documents that are related to one another, to a particular domain”, thus further attesting to the significance of considering provenance as an important part of the metadata endeavour. Figure 7.16 shows the category of metadata provenance.

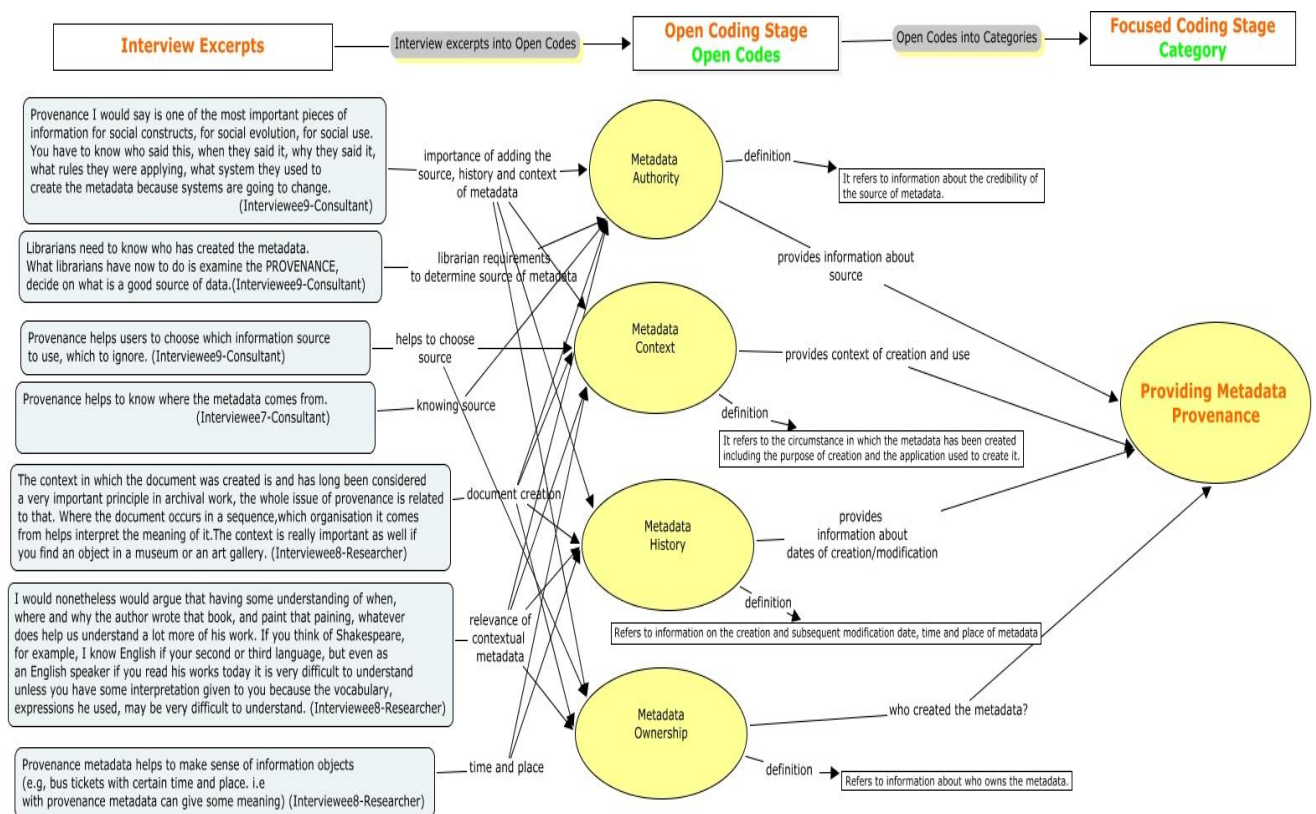


Figure 7.16: The Category of Metadata Provenance

7.4.4. Ensuring Metadata Quality

Whilst recognising the elusiveness and subjectivity of the concept of quality, librarians seem to hold the view that ensuring metadata quality is an important component process of the metadata function, thus indicating the importance of controlling and monitoring metadata records. In this regard, librarians consider their role, as experts, as taking the responsibility for ensuring metadata quality. One interviewee, a librarian, noted that they would not like metadata quality to be compromised and would only permit other metadata creators, apart from librarians, as add-on, rather than as full-fledged metadata entry makers (Study2-Interviewee1). In other words, librarians are convinced that ensuring the reliability and consistency of metadata is their prerogative. Thus, even though libraries continue to purchase their metadata from vendors such as OCLC's WorldCat, librarians maintain that they should check the acquired metadata values against the information objects.

However, the interviewees who were researchers or consultants maintained that the role of ensuring metadata quality could be delegated to other actors, including users. One metadata consultant (Study2-Interviewee11) even contended that the sole prerogative of librarians on metadata quality implies an inhibition to metadata scalability as the time, resources and even expertise required to ensure metadata quality fails to match the quantity of information objects that need to be described. This interviewee believed that metadata records can be continually improved by starting from simple, good enough records, rather than attempting to create a complete and perfect record in one go. Another metadata consultant noted that: "the fact that someone else other than librarians adds metadata doesn't mean that the quality has

gone down. There is some very good information on the Internet and there is some very bad information on the Internet. But it doesn't mean bad information deletes the good information" (Study2-Interviewee7).

The Open Coding data analysis resulted in the identification of the following Open Codes (Figure 7.17):



●	Accuracy
●	Consistency

Figure 7.17: Open Codes relating to Metadata Quality

Another interviewee (Study2-Interviewee12) disputed the view that librarian-created metadata is of high-quality, citing their experience, whilst examining librarian-created records, discovering inconsistent and inaccurate records. The interviewee contended that "There are many different ways to describe a book and the idea that librarians get it right and users get it wrong is a fallacy from the start". Interviewees recognised the significant investment that would be required to ensure metadata quality. To that effect, one interviewee stressed the importance of a minimalist metadata approach, as it is more feasible to control the quality of a few metadata elements rather than rich metadata sets (Study2-Interviewee1). Interviewees, especially practising librarians, are overly wary of delegating control on metadata quality. The Focused Coding of interview data thus resulted in the identification of the Category of "Ensuring Metadata Quality" (see Figure 7.18).

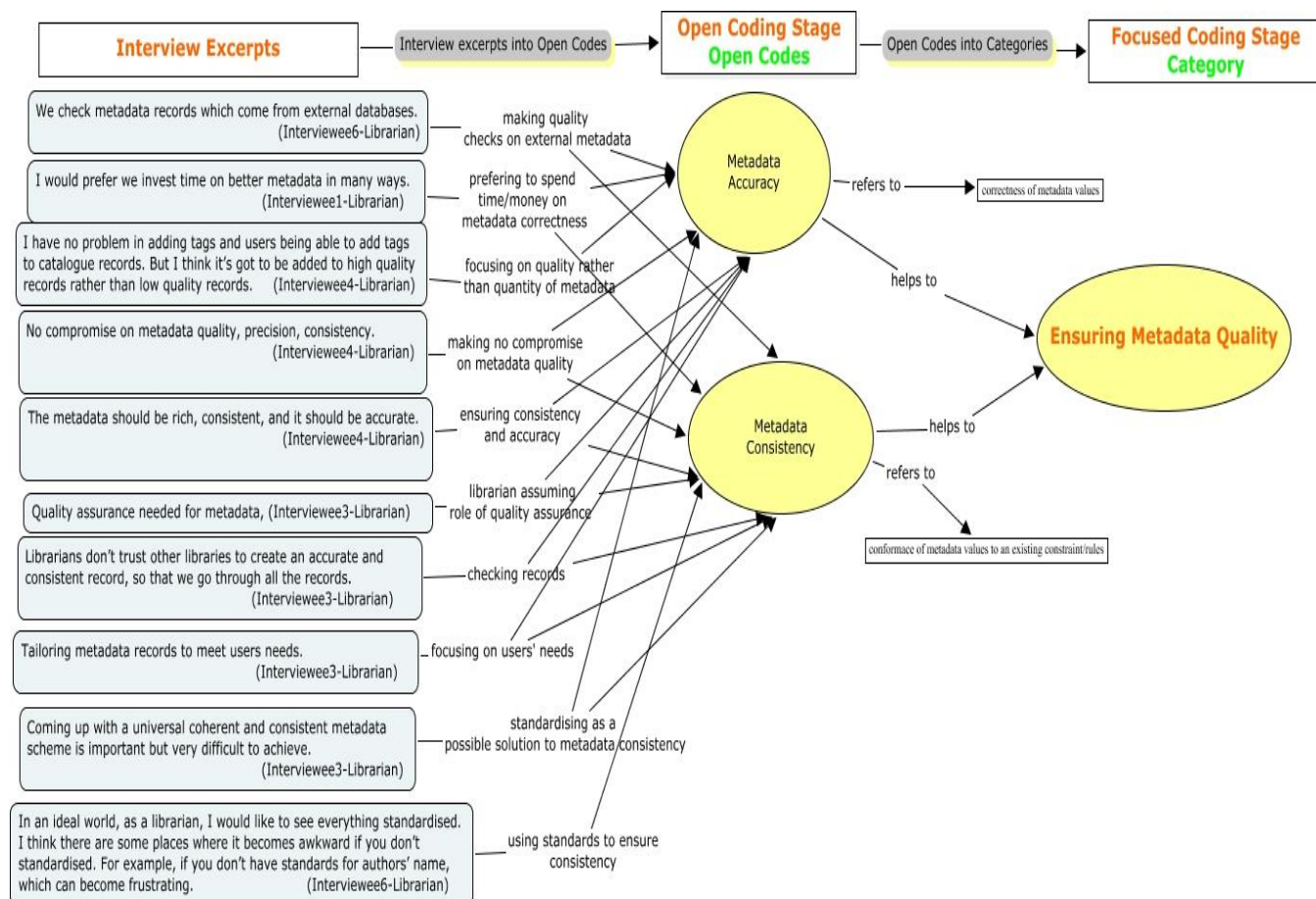


Figure 7.18: The Category of Metadata Quality

7.4.5. Ensuring Metadata Simplicity

Current standards-based metadata approaches, the interviewees stated, have essentially been made to adopt the principle of metadata simplicity or what some interviewees referred to as a minimalist approach to metadata. As the interviewees indicated, metadata simplicity is seen from two vantage points. Firstly, it refers to the number of fields used to describe an information object. Secondly, metadata simplicity indicates metadata usability in relation to the OPAC/discovery interface. One user, for example, described their experience of library metadata thus “the catalogue entry was very basic [as it] adopted a mini-catalogue format” (Study2-Interviewee1). The interviewee further recalled that the minimalist metadata entry affected the metadata project they were involved in, as the previous metadata system had failed to incorporate such critical metadata fields as the ISBN of a book, in order to identify a particular work. “So when the time came and we wanted to upgrade to a commercial system, we had no ISBN entries to search for in the catalogues and find matching records. It was extremely difficult to do.”

Most of the interviewees contrasted MARC with Dublin Core, whereby the former is considered as having many metadata fields (more than 100) whereas the latter only has about 15. Interviewees pointed out the problems associated with metadata simplicity by stating that

there needs to be a balance between what is created and what is presented to users. However, they also acknowledged the associated costs of simplifying the display.

Metadata simplicity has two Open Codes: sufficiency of metadata elements and metadata usability, which both were identified from the Open Coding stage (see Figure 7.19).

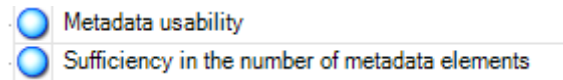


Figure 7.19: Open Codes relating to Metadata Simplicity

The analysis of the Focused Coding stage resulted in the emergence of the Category of “Metadata Simplicity” (see Figure 7.20).

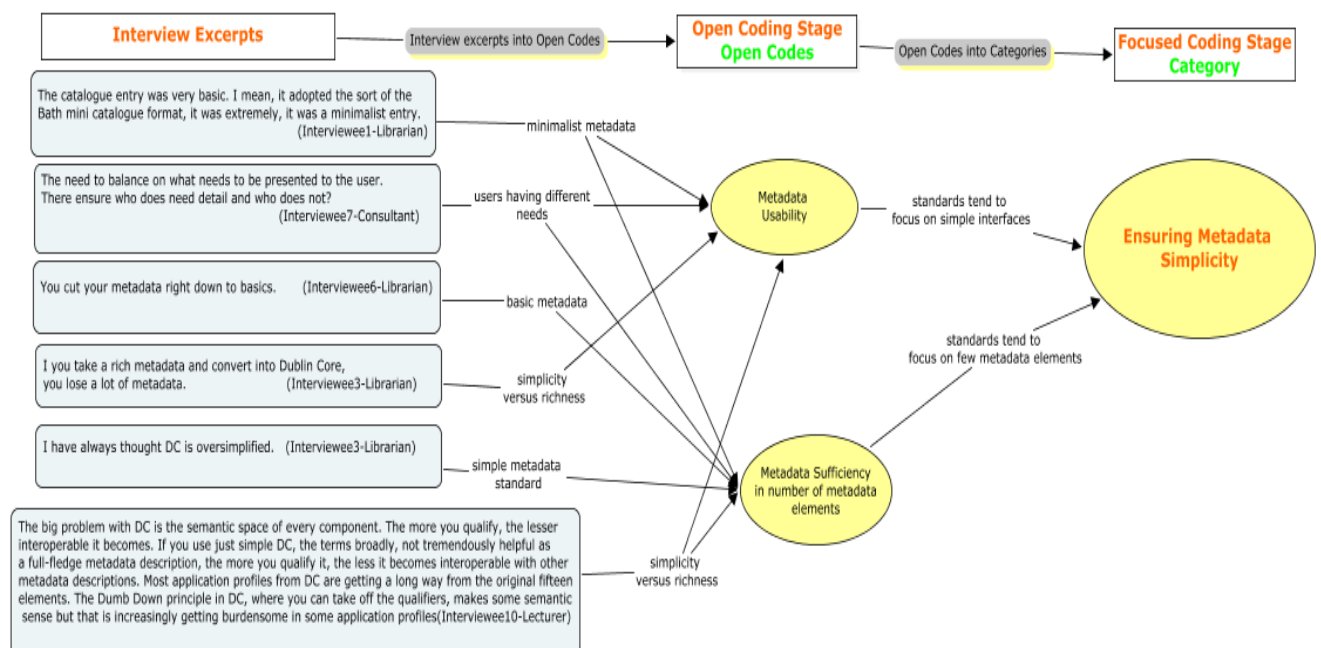


Figure 7.20: Category of Metadata Simplicity

7.4.6. Supporting Metadata Interoperability

In the library and information domain, there are a plethora of metadata standards. The existence of various genres of information objects has necessitated the creation of several metadata standards and record formats. Even within the realm of a single metadata standard such as MARC, there exist dozens of variations and instantiations. Whilst diversity is welcome in view of addressing particular needs, such variations also mean less interoperability. Interoperability refers to the smooth sharing and exchange of metadata records between different standards. Very often, different metadata standards use different metadata elements, naming schemes, encoding rules and controlled vocabularies. In connection with this, one interviewee (Study2-Interviewee10) placed metadata interoperability as one of the most crucial issues that needs due consideration in metadata discussions. Whilst stating that their research was concerned with interoperability, where

emphasis is given to the integration of metadata environments, the interviewee (Study2-Interviewee10) pointed out that interoperability is a multi-layered problem, where it could partly be addressed technically using RDF and XML data formats. The interviewee singled out the METS schema, which mainly used XML, as a tool that could help bridge interoperability issues by allowing the mixing and matching of various metadata elements from various standards and creating what they called an application profile.

Many interviewees noted that, without interoperability, metadata would largely remain in disparate silos. One should also note that metadata interoperability is a major concern of not only libraries but also of other digital information services providers. There are different layers of interoperability; all having important ramifications on the metadata used in libraries (Alemu, Stevens, et al., 2012b). For example, metadata interoperability at the level of metadata fields and constraints, including data formats, implies a technical (syntactic) interoperability and addresses issues at the meaning of the values, whilst metadata richness implies semantic interoperability.

All interviews see Dublin Core (DC) as a standard which is better for interoperability; however, they do not consider it as a fully-fledged metadata standard. One interviewee said that “I have always thought DC is oversimplified. When it was started, it did not have authority control, it is a lot better now but this is not enough for specialist purposes. DC is fine if you get a website. But if you get a complicated learning object, DC fails to scale. It does not allow you to describe in greater granular detail” (Study2-Interviewee3). Another interviewee (Study2-Interviewee10) also noted that the unqualified DC metadata element set offers better metadata interoperability, whilst concurring that it has limited use as a full-fledge standard. The interviewee indicated their wish to see a scalable metadata standard that supports metadata richness whilst at the same time supporting interoperability. Because as the interviewee noted, at present even the qualified Dublin Core element set and its various instantiations (such as using application profiles) suffers from interoperability challenges. Thus, as more metadata elements are added to a metadata schema, the issues of interoperability should be considered.

Cross-walking (see Figure 7.21) between standards is considered as an important and common way of achieving metadata interoperability. One interviewee recalled their experience of cross walking (mapping) between records from simple metadata standards, such as Dublin Core, and more complex ones, such as MARC. However, the interviewee noted the limitations of such cross-mapping between standards:

“I would always be hesitant, if you have got a rich set of metadata that you are not tempted to simplify so much that you lose some of that richness. But it is still important to make it available in a format that is consistently cross searchable. The point is that you can always express rich data in a simple way, but you can’t express simple data in a very rich way because you’ve lost the mark-up, you’ve lost all that description. So, if you got systems that have got rich data in them, you might as well keep them that way, as you can always express it in a variety of different ways” (Study2-Interviewee4).

Another interviewee believed that adopting a single international standard would be the best means of ensuring interoperability. However, the interviewee (Study2-Interviewee3) noted the numerous challenges to coming to an agreement on using such a global standard that should have a comprehensive number of metadata elements. In addition, the interviewee pointed out the challenges associated with the conversion of legacy metadata to this new standard.

According to one interviewee, any changes to standards-based metadata have always been done from top down whereby each activity has to be coordinated in such a manner that adequate time has to be given to libraries and vendors, so as to implement these changes successfully. The challenges in this regard are not only technical and financial, but mainly social and political. These procedures take a very long period of time to effect any meaningful changes towards enriching metadata. However, another interviewee argued that a decentralised approach to metadata provides better flexibility when adapting to changing needs. In order to manage metadata changes at a decentralised level, effective coordination is essential, in order to make various metadata repositories interoperable. The interviewees emphasised the importance of metadata cross-mapping, thus the identification of “Cross-walking” as an Open Code during the Open Coding stage. This in turn led to the emergence of “Ensuring Metadata Interoperability” as an emerging Category (see Figure 7.21).

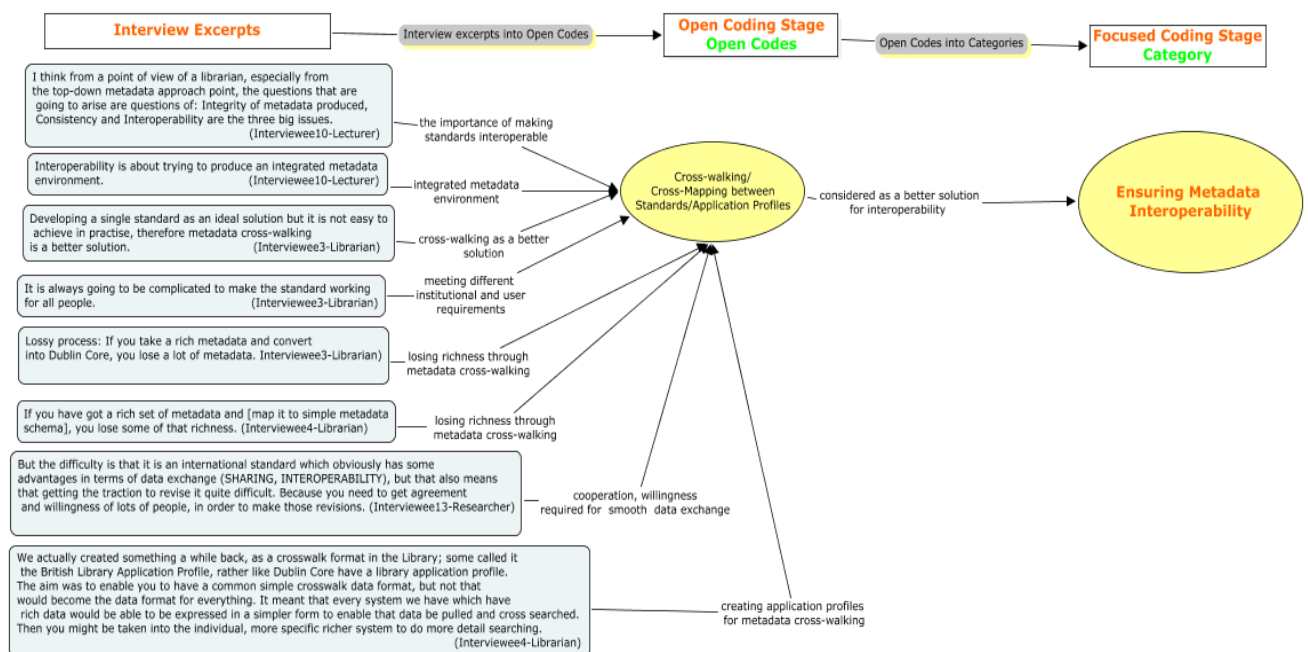


Figure 7.21: Category of Metadata Interoperability

7.5. Summary of Categories of Standards-based Metadata Approaches

As discussed in detail from Sections 7.4.1 to 7.4.6, six Categories emerged from the Focused Coding of interview data (see Figure 7.22). The categories included: providing metadata structure, granularity, provenance, quality, simplicity and interoperability. The interviewees considered these categories as the major components of metadata standards that need to be considered.

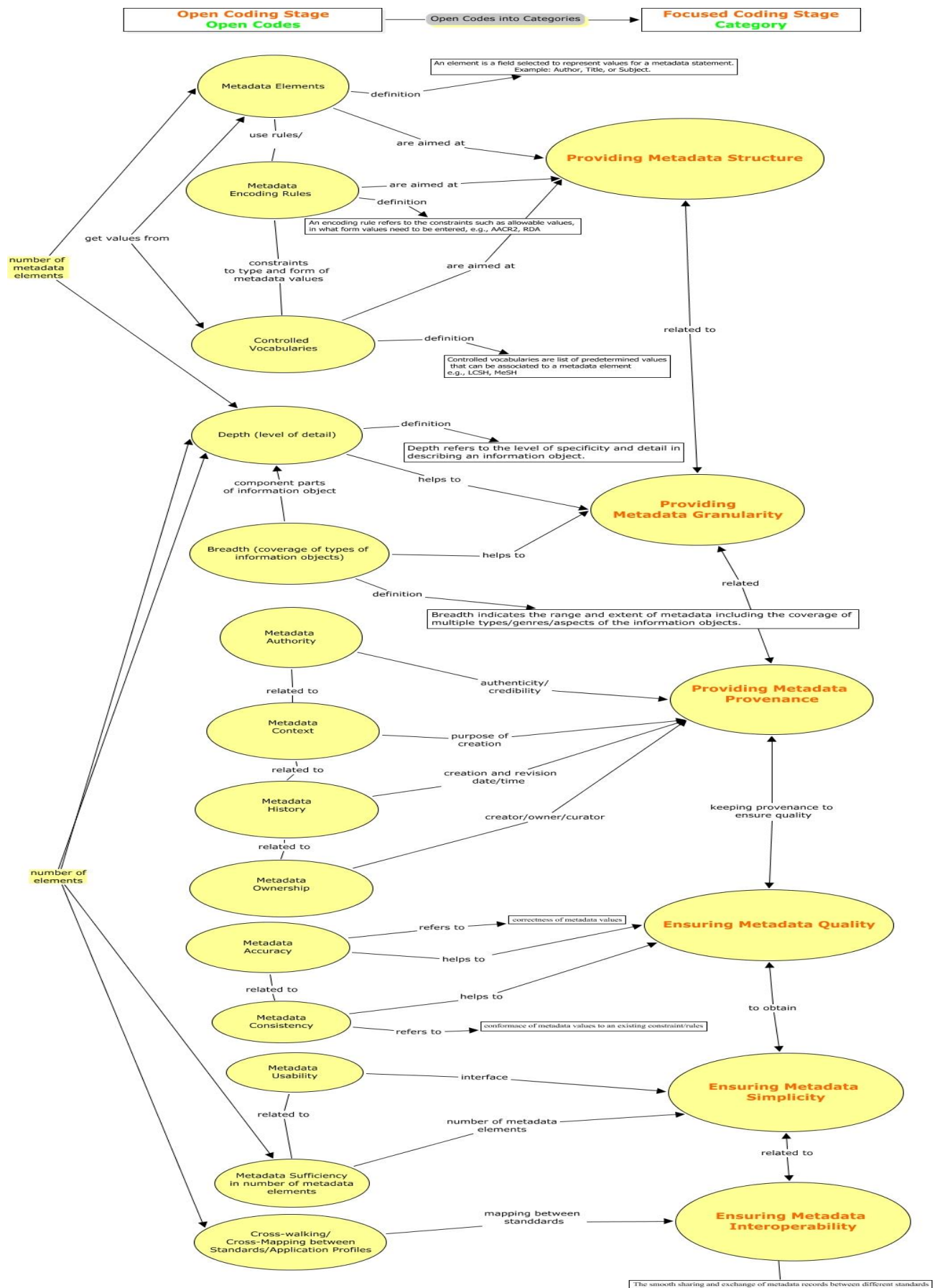


Figure 7.22: Summary of Categories of Standards-based Metadata Approaches

7.6. Challenges of Standards-based Metadata Approaches

In as much as standardised metadata approaches are beneficial for ensuring structure, granularity, provenance, quality, simplicity and interoperability, as the interviews revealed, this doesn't come without limitations. The following are some of the major challenges (limitations) of standards-based metadata approaches that emerged from the interview data during the Open and Focused Coding: rigid structures, metadata silos, metadata duplication, slowness in accommodating changing needs, and lack of engagement with users.

7.6.1. Structural Issues

One interviewee indicated that their library has got “a lot of data in the MARC format. So we have to work out with the transitioning, in order to maintain the richness of what we have and preserve the investment we have put into that legacy data, in order to blend it consistently with the new generation of data” (Study2-Interviewee4).

In connection to this, another interviewee identified two major challenges:

“Firstly, we already have hundreds of millions of data in these formats and data records. Secondly, the library systems that libraries buy from commercial vendors, that primarily provide systems only to libraries; these library systems only know these types of data. Because libraries share their data, that is when a book is published, the national library of that country generally creates a metadata and all the other libraries take copies of that metadata and put that in their systems. That is quite difficult. Even making small changes in the data standard, has a very widespread implication” (Study2-Interviewee7).

The interviewees recognised that current metadata functions are heavily tied to the use of existing standards and any attempts to change should be cognisant of the resource implications. One of the big challenges for the move from existing to new metadata formats is the size of the existing legacy metadata. In connection with this, one interviewee said that:

“Nobody knows how many MARC records exist but we know that there are lots of them. For example, OCLC's WorldCat has about 200 million metadata records, and WorldCat constitutes less than 10% of the metadata records in the world, [...] with that kind of bulk, it is very difficult to change systems from the point of the library management system, vendors and developers. [In addition] library systems vendors operate with very small profit margins, because the market is very restricted. Because of this small profit margin, they are unable to afford to experiment very much” (Study2-Interviewee9).

Another issue raised is the problem in relation to controlled vocabularies and metadata values in general and the absence of a global identification mechanism for data. Most of the efforts in this direction are piecemeal at best, and hence, a consistent, agreed-upon system of a global naming space has yet to emerge (for discussions on unique identification mechanisms see 10.4.2). Interviewees noted the importance of distinguishing between entities and their labels. A single concept, as an entity, can be uniquely identified and any other concept can be

attached to it, without any linguistic limitations. Once such limitations are made superfluous, the language of the user, author or metadata expert can enrich the said information object (see 10.4.3).

Even though current library metadata principles in theory support the importance of having structures and granularity, existing formats such as MARC have been severely limited at providing the atomicity and detail required to describe resources at a granular level. Consequently, at the moment most library metadata, as the interviewees indicated, is not machine process-able.

Based on the interview data, the Open Coding analysis stage thus resulted in the identification of an Open Code called “Structural Issues”.

7.6.2. Metadata Silos

As the data analysis from the interviewees showed, within the remit of libraries, a plethora of standards exists. The standards which are in use in libraries are not interoperable with standards in other domains. The metadata, thus created using library standards remains to be used solely by libraries. The interviewees indicated that the resultant metadata from standards-based metadata approaches result, very often, in metadata that can only be used in libraries. Besides, the formats are not designed in such way that they can be mixed and matched with other metadata, thereby resulting in what the interviewees called metadata silos. Interviewees believed that library data should not be “locked up in various silo systems” (Study2-Interviewee12). These challenges should be addressed in view of current and emerging technologies, as well as users’ and institutional requirements. One of the solutions would be investigating the importance of metadata openness (see Chapter Eleven). In general, the Open Coding stage of interview data helped to identify “Metadata Silos” as an Open Code.

7.6.3. Metadata Duplication

Contemporary metadata structures and formats were designed in the 1960s where data processing technologies and formats were still primitive. A case in point is the MARC format, which most interviewees concurred that its structure does not obviate the creation of duplicated values and records. One interviewee (Study2-Interviewee14) noted that “cataloguers sometimes create the same bibliographic record [multiple times]”. This is partly because, as the interviewees indicated that cataloguing principles such as AACR2 and formats such as MARC do not necessitate the creation of machine process-able and re-usable metadata, thus librarians could not easily work through to obviate the duplication of data within even a single description of an information object. One metadata consultant (Study2-Interviewee12) concurred saying that “libraries should not re-invent the wheel” by adding new descriptions each time metadata is created for a book. Expressing their views on the MARC standard, the most widely used metadata standard, the interviewees singled out the problems of metadata record duplication, lack of scalability and format obsolescence as the main challenges. Another interviewee stated that:

“If you look at the way a librarian catalogues a book in MARC, he/she creates an individual record for a book that is in their hand. So if you catalogue, let us say Hamlet by William Shakespeare, you get a catalogue card, a typical MARC record that has a lot of duplicate information in it, like who the author was, when he was born, when he died, publisher and various pieces like that. If you then catalogue another book by William Shakespeare, again you would need to duplicate it all in a similar fashion as well as the book itself, the joining of the author and title” (Study2-Interviewee11).

Thus, from the Open Coding interview data analysis stage, the Open Code of “Metadata Duplication” was identified as an outstanding challenge of existing metadata standards.

7.6.4. Slowness in Accommodating and Adopting the Changing Needs of Users

Standards-based approaches are slow in identifying, accommodating and adopting the changing needs of users on several fronts. One of the concerns mentioned by interviewees was the incongruity between the terminologies found in structured subject headings, on one hand, and those users employed when searching library databases, on the other. One interviewee asserted that the controlled vocabularies used in library databases are updated rather infrequently (Study2-Interviewee1). According to the same interviewee, the National Library of Medicine’s Medical Subject Headings (MeSH) is updated every five years (Study2-Interviewee1). Both interviewees indicated that the inclusion of new and emerging vocabularies from particular disciplines as well as incorporating the terminologies that are likely to be used by users of these systems remain important challenges.

Furthermore, interviewees indicated that the MARC standard has become anachronistic. One interviewee ascribed some of its current limitations to the cataloguing policies used in libraries. One interviewee (Study2-Interviewee5) also attributed the problems of OPACs/discovery interfaces to that of the anachronistic nature of library standards. In addition, the interviewee believed that librarians and cataloguers at present focus, perhaps unnecessarily, on cataloguing policies and rules, but from a resource discovery point of view, such rules are confusing to the user. For example, as some of the interviewees noted, at present the MARC record format does not support links and displays that meet the requirements of users. The format is said to focus too much on physical description and less on the essence of information objects.

From the interview discussion above and from the Open Coding stage of interview data analysis, the issue of “Slowness in Accommodating to Changing Users’ Needs” was identified as an important challenge.

7.6.5. Lack of Engagement with Users

One interviewee pointed out the need for change in some of the standards-based approaches such as the use of the MARC standard format. New metadata approaches should be embraced in view of improving the user experience. The same interviewee acknowledged that librarians are well aware of technological developments and also the need for new metadata approaches

but they are still tied to their old formats and approaches that are well established in libraries. One interviewee acknowledged the enthusiasm of librarians towards engaging with their users but this has not been translated into practise, such as in making user-contributed metadata a core part of the catalogue and search interface. Another interviewee noted:

“When you come from a rigid, structured data model, and that is what librarians are dealing with, it is not easy to break that model by putting data derived from a model designed for a different purpose, data with less quality. This is where the user-driven tagging world clashes with the hierarchical subject world of libraries” (Study2-Interviewee11).

The interviewee noted that they see no future for MARC as a reliable standard but admitted that the process of converting MARC records would require substantial investment. There is consensus among interviewees that the uses of terminologies in metadata are context dependent. Some of the challenges mentioned in this regard include broader/narrower terms, synonym and related terms, differences emanating from British and United States usage of varying terms and spellings, and scientific versus day-to-day uses of language. As one interviewee noted, one of the challenges librarians face is whether to use words from the literary warrant, that is as they appear on book titles and covers or to use concepts that represent the contents of an information object.

Thus, the Open Coding stage resulted in the identification of the Open Code called “Lack of Engagement with Users”, indicating the importance of involving users in metadata creation and utilisation.

7.6.6. Summary of the Challenges of Standards-based Metadata Approaches

As discussed from Sections 7.6.1 to 7.6.5, the Open Coding of interview data resulted in the identification of Open Codes: Structural Issues, Metadata Silos, Metadata Duplication, Slowness in Accommodating the Changing Needs of Users and Lack of Engagement with Users. These Open Codes were considered as the major issues mentioned by interviewees. Thus, during the Focused Coding stage, these Open Codes were brought into one Category called “Challenges of Standards-based Metadata Approaches” (see Figure 7.23).

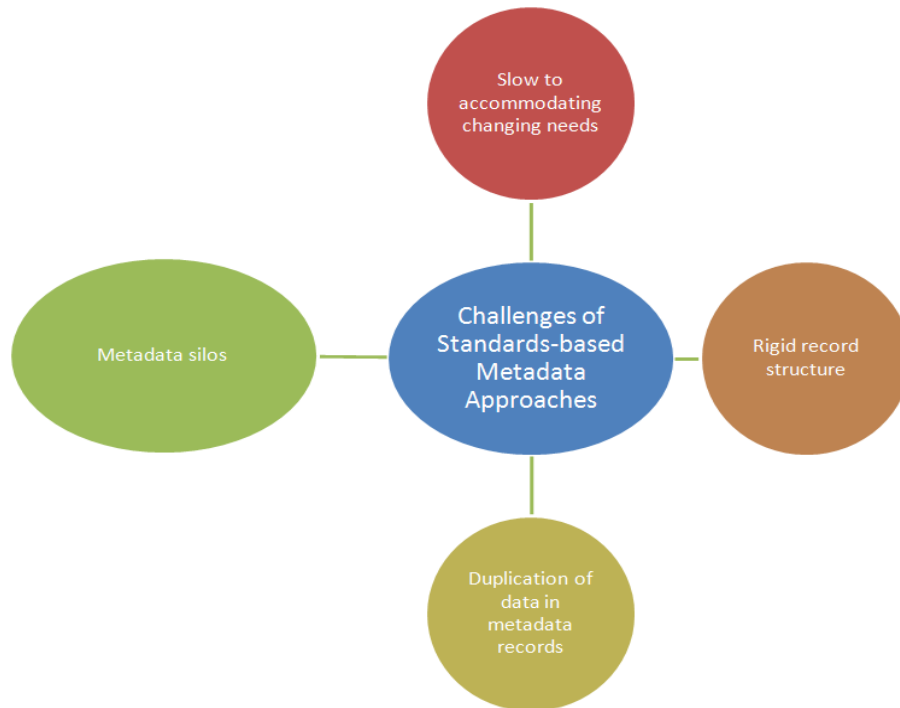


Figure 7.23: Summary of the Challenges of Standards-based Metadata Approaches

7.7. Conclusions

Basic or core metadata, which is further conceptualised in this research as a priori metadata (see 7.3.3), is created by metadata experts using standards-based metadata approaches. When it came to the issue of bibliographic metadata, as the responses from the interviews revealed, libraries have an enduring history of using standards in their metadata functions, and indicated that these standards are tightly meshed into the operations of librarians. As the data analysis indicated, whilst current standards-based metadata approaches play an important role in providing metadata structure (7.4.1), granularity (7.4.2), provenance (7.4.3), quality (7.4.4), simplicity (7.4.5) and interoperability (7.4.6) through the enforcement of pre-defined and pre-filtered metadata schemes that include metadata elements, controlled vocabularies and encoding rules (7.4.1), its scalability to enrich the ever increasing collections in libraries with rich metadata is considered limited. However, whilst standards in principle aim to provide structure, granularity and interoperability, contemporary formats fail to materialise them in practise. Thus, currently, standards-based approaches are confronted with challenges such as non-machine-process-ability, metadata silos (lack of interoperability with external non-library metadata), metadata duplication, slowness in accommodating users' needs in terms of user vocabularies (search terms) and lack of engagement with users (7.6). Some of these challenges can be addressed through technical means, such as changing library metadata formats from MARC to web-compatible formats, others are conceptual and yet others are social and institutional. For instance, the issue of whether users can be involved in the creation and management of metadata can be attributed to social and institutional challenges. To this end, the following chapter (Chapter Eight) looks into the perspectives of LIS professionals on the role of socially-constructed metadata approaches.

Chapter Eight: Study Two - Open and Focused Coding of the Perspectives of LIS Professionals on Socially-Constructed Metadata Approaches

8.1. Chapter Overview

As discussed in Chapter Six (Proof-of-Concept, Study One), an analysis of the perspectives of eleven LIS post-graduates and academics identified the potential role of the Web 2.0 paradigm for metadata functions. In particular, in Chapter Six, socially-constructed metadata was identified as an emerging approach with promise for metadata creation and utilisation (see 6.6; 6.12). The analysis for this Chapter followed a similar pattern to the one used in Chapter Seven (7.1 & 7.2), whereby the **Open Codes** from the Open Coding stage were subsumed into broader themes, called **Categories**, during **Focused Coding**. Similarly, the interview data that dealt with socially-constructed metadata approaches was also coded using the NVivo 9 software.

This chapter thus focuses on the categories that emerged from the Focused Coding stage relating to socially-constructed metadata. The phrases socially-constructed metadata and social space of metadata are used synonymously, both indicating the ability of users to co-create metadata in an online environment. From the Focused Coding, several categories emerged including post-hoc metadata creation (8.2), platform for metadata creation (8.4.1), users as proactive metadata co-creators (8.4.2), metadata diversity (8.4.3), metadata scalability (8.4.4) and collective metadata intelligence (8.4.5). In addition, categories including motivation (8.5) and quality control (8.6) for socially-constructed metadata also emerged from the data analysis. These categories were considered representative of the interviewees' responses and are used to structure the presentation of this analysis.

8.2. Post-Hoc Metadata Creation

Chapter Seven (7.3.3) discusses a priori metadata creation where metadata experts create what the interviewees considered as basic or core metadata using standards-based metadata approaches. Conversely, as indicated by interviewees, if given the opportunity, users would also create metadata by way of add-on or supplementary metadata. Whilst, the purpose of a priori metadata, among other things, is to make information objects findable and discoverable by users, socially-constructed metadata could enhance it. As the data analysis shows, instead of passively using metadata to find and discover information objects, users can in fact assume the role of adding metadata through various mechanisms as afforded by the Web 2.0 paradigm. For the sake of comparison to a priori metadata (before publication of an information object), the metadata created by users can be referred to as post-hoc metadata (after publication of an information object). It is important to note that the interviewees did not specifically use the phrase post-hoc metadata, however, as grounded theory permits and even encourages the use of phrases that have interesting resonance (grab), the phrase post-hoc metadata is considered more explanatory. Figure 8.1 shows the various names interviewees ascribed to the phenomenon of socially-constructed metadata.

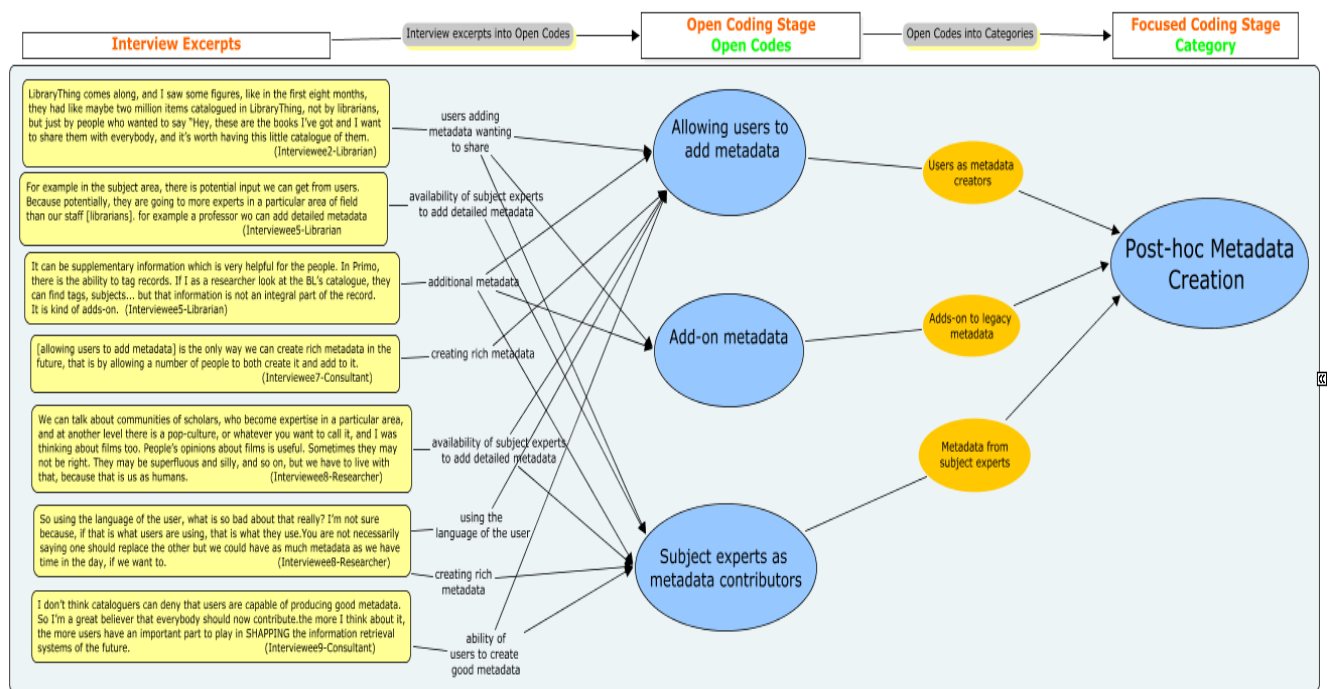


Figure 8.1: The Category of Post-Hoc Metadata Creation

As Figure 8.2 shows post-hoc metadata is an instantiation of the broader approach of socially-constructed metadata. It is worth noting that in order for users to add post-hoc metadata to information objects, expert-created a priori metadata (7.3.3) is crucial, for, without it, users would not be able to access the information object in the first place. Once users find/discover information objects, they could then contribute metadata in the form of tags, ratings, reviews or recommendations.

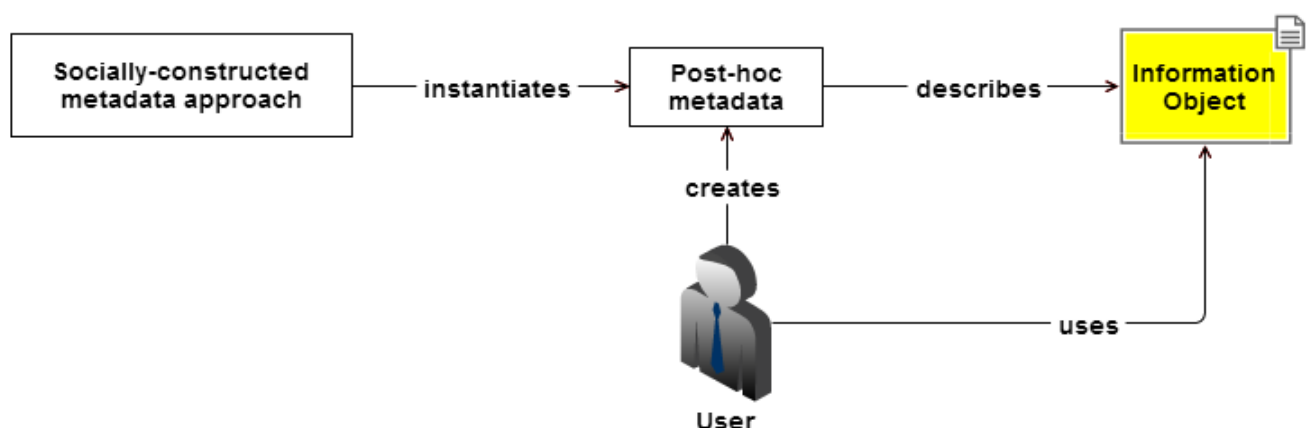


Figure 8.2: Socially-Constructed Metadata Approach Instantiating Post-Hoc Metadata

As the main scope of this section is dealing with socially-constructed metadata approach, the following sections focus on analysing the interview data pertaining to its potential roles, benefits and challenges.

8.3. The Potential Benefit of Involving Users

Whilst interviewees' views coalesce on their being fully aware of socially-constructed metadata approaches, individual responses on how these approaches affect library metadata functions and how it benefits users varied. It emerged from the analysis that the use of Web 2.0 technologies in library use, especially in academic libraries, is in its infancy. Interviewees, for example, acknowledged the absence of metadata models and approaches that demonstrated how socially-constructed metadata could be used in libraries. As one interviewee noted "the challenge is to design a system that can cope with [socially-constructed] metadata and rationalise it" (Study2-Interviewee10). This interviewee's view of rationalisation is in relation to integrating user metadata with legacy metadata, which the same interviewee admitted is quite challenging at present.

Another interviewee believed that the problem is associated with the novelty of Web 2.0 paradigm in general, which can be attributed to the absence of clarity of its very definition. As the interviewee further asserted, "Web 2.0 is a label looking for a description" (Study2-Interviewee11). Unlike standards (top down), being a bottom-up approach, socially-constructed (Web 2.0) metadata approaches may take time until clear definitions are formulated or principles and guidelines established. As the interviewees indicated, whilst there is no absence of enthusiasm about the Web 2.0 paradigm, libraries are still grappling with the issue of fully embracing it in their major functions. Even more challenging is bound to be the prospect of integrating Web 2.0 with standards-based metadata approaches.

Thus, many interviewees maintained that socially-constructed metadata approaches should be given due consideration and that they anticipated the potential of it being integrated along with legacy standards-based metadata so that it helps improve the findability and discoverability of information objects. They, nonetheless, suggested the need for predicating the decisions of incorporating such metadata on sound metadata principles. The interviewees further highlighted the need for exerting the requisite social and technological efforts in the formulation of viable principles and approaches to effectively utilise the Web 2.0 paradigm in library metadata functions.

However, the interviewees did note their belief in the potential of users being involved in the co-creation of metadata. One interviewee (Study2-Interviewee11) for instance foresaw the possibility of socially-constructed metadata becoming a major component of metadata, from which the library catalogue and search interfaces could benefit. At present, however, most interviewees considered socially-constructed metadata as an emerging, but yet unsettled approach.

Whereas another interviewee (Study2-Interviewee18) emphasised the need to differentiate the cases where socially-constructed metadata may help and where it does not. The interviewee noted that they would not want to see users editing librarian created metadata.

Instead, if needed, they suggested the use of a separate metadata field alongside the standardised fields so as to then let users enter metadata values. It is worth noting that one interviewee (Study2-Interviewee1), in particular, stressed their concern on the potential of involving users in the creation of metadata, indicating that not all interviewees were fully convinced of the future of socially-constructed metadata approaches. However, this interviewee did not want to sound as totally objecting to the notion of socially-constructed metadata, but expressed their concern that involving users in the creation of metadata might degrade the quality of the metadata, and hence lower the institution's (library's) reputation as a source of credible and trustworthy metadata. However, the said interviewee acknowledged the possibility of involving users in the future, by enabling them to provide add-on metadata, on top of the metadata created by metadata experts; however, they maintained that such metadata should be kept separate from standards-based, and hence expert-created metadata. The interviewee went on to add that, even when users are allowed to create add-on metadata, their input should be checked, vetted and controlled by experts, such as librarians, so as to ensure its veracity and accuracy before it is released for general use. The interviewee further expressed their concern regarding the availability of resources for paying librarians to quality-check and police socially-constructed metadata. This view was also shared by another interviewee (Study2-Interviewee18) who cited resource constraints potentially limiting librarians to spend their time policing user contributed metadata. In contrast, three interviewees (Study2-Interviewee2; Study2-Interviewee3; & Study2-Interviewee9) alluded to the concept of the network effect (8.4.5.5) and the "Wisdom of Crowds" (8.4.5.6), where the value of socially-constructed metadata lies in metadata being aggregated (8.4.5.3).

Put another way, whilst personal metadata may have maximum benefit to the person who created it, it is the aggregation of such individualised contributions that provides the social space of information objects.

"I've had a little catalogue of my books at home, ever since I was a kid. That's perhaps sad and geeky, and maybe it is not practiced widely and yet LibraryThing comes along, and I saw some figures, like in the first eight months, they had about two million items catalogued, not by librarians but just by people who wanted to say 'Hey, these are the list of books I've got and I want to share them with everybody'" (Study2-Interviewee2).

The interviewee added that the value of social metadata increases with scale, hence its benefit only becomes apparent as more users start conversations about an information object. Such conversations may take the form of reviews, tags and/or recommendations. Interviewees also pointed out the problematic nature of predetermining the value of individual metadata contributions, as it would be difficult to anticipate the relevancy of metadata prior to its use. Thus, it is metadata usage that provides an important clue to the relevance of a particular metadata. Some metadata, even though seeming very idiosyncratic may be of some value to someone, which can only become apparent when it is allowed to be used in a social environment. In connection with this, one interviewee stated:

“My dad catalogues his books. He’s got thousands of books in his attic, where he works in a little office. And we set up a little catalogue for him so he can put the title, where they are, and he can find them much more easily and he knows that he’s got a particular book. He knows he’s got the book somewhere, but can’t find it easily. And when we set up the fields that he was going to put in such information, he decided that he would like to be able to put the colour of the cover of the book into the records, so that he can think ‘I am pretty sure that it’s got an orange cover’. And it’ll help him when he looks it up in the catalogue to be able to find the book on the shelves, and again, librarians don’t typically catalogue books by colour” (Study2-Interviewee2).

In view of socially-constructed metadata thus, pre-filtering metadata values becomes challenging or even unwarranted, as it is difficult to anticipate future use of metadata. This can be put in sharp contrast to expert-created metadata which anticipates its future use by others, thus the need for constructing a pre-determined list of metadata fields (a priori metadata, see 7.3.3) and assigning them metadata values derived from authoritative subject headings and authority lists (see also 7.4.1). Since users predominately add metadata, such as tags, for their own retrieval purposes, it brings implications on whether such metadata should be edited or removed and, if so, on what basis.

Whilst standards-based metadata approaches, on the basis of such principles as sufficiency and necessity (see 2.6), mandate that only metadata that is deemed significant should be maintained and utilised, socially-constructed metadata approaches tend to obviate some of these limitations. Hence new metadata principles and new solutions to the issues that socially-constructed metadata pose should be investigated.

It is important to note, from the above narrative of the interviewee, that users’ queries may not always necessarily correspond to formal library cataloguing. The emphasis here is not so much about whether the colour of a book should be an important metadata field; however, it suggests the importance of reflecting users’ terminologies in library metadata. The interviewees further indicated that socially-constructed metadata approaches inherently incorporate the potential to make the library an interesting place where conversations and interactions can happen over social networks. One interviewee (Study2-Interviewee2) thought that “if the library catalogue was a much more desirable place to be, because it was not simply [the] boring, traditional type library catalogue, but has these social elements, not only could you add the metadata but you could also perhaps engage in conversations, may be like the ones on Facebook likes, games or something that is connected to what is going on or just enable the social interaction”. Another interviewee (Study2-Interviewee9) remarked on the importance of tapping into the social space of information objects through socially-constructed metadata. Valuable insight can be gleaned from such interactions and conversations around information objects, which add weight to the case for socially-constructed metadata approaches.

In connection with this, David Weinberger (Weinberger, personal communication, December 6, 2012) underscored that the issue of involving users in metadata creation depends upon the purpose libraries want to achieve. He argued that:

“If you are trying to help people find things, I think you do better by accepting a messy world rather than a well ordered world. You can train people to think in terms of your taxonomy which is expensive and inefficient or you can do that and create a labour intensive system of librarians who can interpret the taxonomy for you or interpret the requests and put them into the controlled system of the taxonomy”.

As Weinberger explained, taxonomies are good at precision but poor at recall; the opposite is true for folksonomies. Weinberger indicated that the standards-based metadata approach works well for small size library collections but with the enormity of information objects that abound in libraries at this point in time, it poses scalability problems, thus re-orienting current metadata approaches is paramount. Whilst it has significant benefits to involve users in metadata creation and thus utilise the metadata in an effective manner, failing to do this has broader consequences:

“Well, one approach would be to not allow any of that metadata and to throw it out. That is a tremendous loss and it will cause most users to seek other alternatives to the library because the library is not returning enough information. Certainly when there is a need for a great deal of precision and reliable sources, the library will continue to have its place. Another alternative is to accept this new metadata and to deal with it” (Weinberger, personal communication, December 6, 2012).

Once a decision is made to consider socially-constructed metadata, the technical challenges of utilising the metadata can be then addressed. As Weinberger indicated, the solutions could be algorithmic, crowd sourcing (folksonomic) and/or metadata-about-metadata approaches. The latter approach refers to providing additional metadata and making the source and context of metadata explicit to the user, thus users are kept informed about the authority and veracity of the information, indicating thus the importance of provenance information for socially-constructed metadata (7.4.3; 8.4.5.8).

In summary, socially-constructed metadata is considered a relatively new phenomenon for libraries, and thus has not made remarkable inroads into the actual metadata creation and utilisation functions. Whilst there is no lack of enthusiasm for socially-constructed metadata approaches, what is perhaps lacking is a careful investigation and exploration of the potential roles this emerging approach provides, including identification of the motivations that drive users, so as to effectively harness it. As several interviewees indicated, there are several interrelated issues that should be addressed and carefully examined before these approaches are put into practise. Interviewees indicated several challenges that need to be addressed in relation to the potential of involving users as metadata creators. Focused Coding of interviewee data helped identify the following **Categories** of socially-constructed metadata approaches. These Categories are used to structure the discussions in subsequent sections.

8.4. Open and Focused Coding Relating to Socially-Constructed Metadata Approaches

The Open Coding stage of interview data analysis resulted in a number of Open Codes which were, during Focused Coding, subsumed under a higher hierarchy called Categories. In order

to add readability to the discussion, the Open Codes are subsumed under their broader Category. For example: instead of first presenting and discussing the Open Code called ‘Tagging’, it is presented under its higher Category, i.e. The Category of “**Platform for Proactive Metadata Co-creation**”.

The Focused Coding of the data analysis resulted in the emergence of the following Categories which are considered essential in the consideration of socially-constructed metadata approaches (Figure 8.3):

- The Category of Platform for proactive metadata co-creation (8.4.1)
- The Category of Users as proactive metadata co-creators (8.4.2)
- The Category of Metadata diversity through social metadata (8.4.3)
- The Category of Metadata scalability (8.4.4)
- The Category of Collective metadata intelligence (8.4.5)

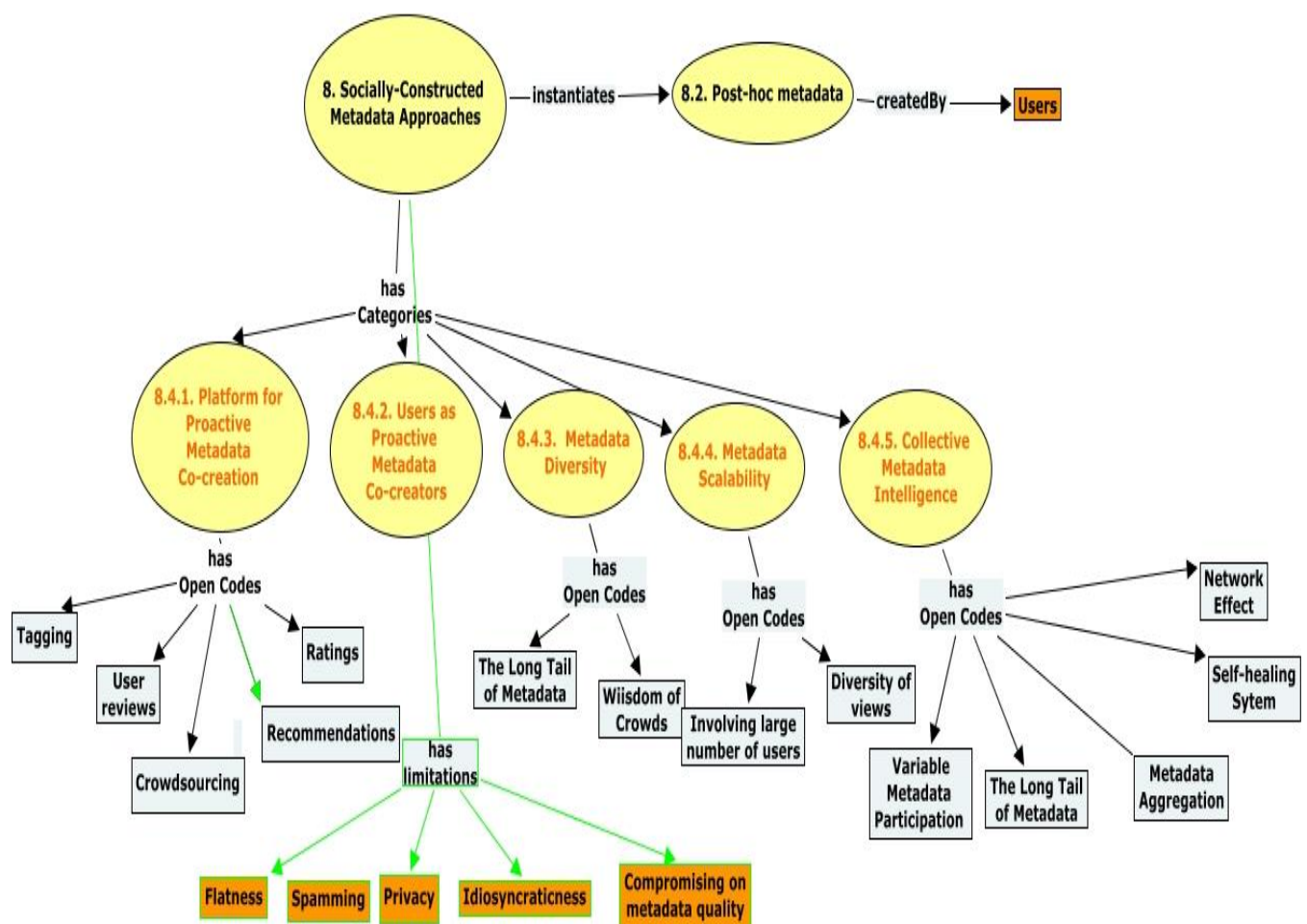


Figure 8.3: Categories of Socially-Constructed Metadata Approaches

8.4.1. Platform for Proactive Metadata Co-creation

During Open Coding analysis of interview data, the following **Open Codes** were identified pertaining to the technological platform for proactive metadata co-creation:

- **Tagging** (8.4.1.1)
- **User Reviews** (8.4.1.2)
- **Ratings** (8.4.1.3)
- **Recommendation Systems** (8.4.1.4)
- **Metadata Crowd Sourcing** (8.4.1.5)

Subsequently, during Focused Coding, these Open Codes were put under the Category of “**Platform for Proactive Metadata Co-creation**”. Interview excerpts are used to support the discussions under each section. But the following is an after-the-fact structuring of the discussions of the Open Codes.

8.4.1.1. Tagging

During Open Coding, ‘Tagging’ was identified and is thus used to present and discuss the perspectives of interviewees associated to the process of attaching labels or keywords to an information object. As the interview data showed, upon discovering an information object, tagging allows users to mark it with one or more terms. Interviewees identified tagging as the most dominant manifestation of socially-constructed metadata approaches. One interviewee described tagging as an “informal, organic and user-driven” approach to metadata (Study2-Interviewee2). As the said interviewee admitted, they are familiar with tagging tools, such as the ones implemented in Flickr (a social photo sharing website), Delicious (a social bookmarking website) and LibraryThing (a social cataloguing website). The interviewee further noted that they found the process both interesting and handy for cataloguing their personal collections of books, while sharing this online was fun, interesting and useful. As the interviewees pointed out, some of the unique characteristics of tagging include the fact that it is created by users, that the metadata assignment process happens post-hoc (i.e. after it has been published), that the assignment of tags obviates any pre-existing classification (taxonomy) or semantic structure and that it follows no cataloguing/metadata rules in the assignment.

One interviewee (Study2-Interviewee3) indicated that “people tag resources from their own perspectives”, reflecting their own personal understanding of the information object and their preference for terminologies. In this regard, as three interviewees (Study2-Interviewee2; Study2-Interviewee12; & Study2-Interviewee21) identified, the primary purpose of tagging is users wanting to re-find (retrieve) information objects at a later time. One interviewee (Study2-Interviewee14) stated that tagging provides various avenues of access to the same information object. The same information object can be tagged variously using a plethora of

terms, and these tags could be used as entry points to the object. The interviewee compared this phenomenon with the standards-based approach of current classification in physical libraries, where an item can only be shelved in one particular location. Whereas, as digital libraries obviated the constraint of location, an information object can be categorised and placed in various places (Study2-Interviewee14). Each user tag, as the interviewee pointed out, enables a new metadata environment where there could be infinite possibilities to access an information object. However, not everyone agrees with the use of all tags in organising information objects. One interviewee (Study2-Interviewee18) believed that ‘not all tags are created equal’, thus pointing to the importance of choosing the most pertinent ones. The most popular tags (those terms which are frequently used by the majority) should be perhaps included within the librarian created metadata.

User tagging is a process that essentially occurs post-hoc, i.e. an object must be findable before tagging can occur. However, the platform of the library system should be designed in such a way that users can attach tags to information objects and should also be able to support searching and finding of those information objects using the allocated tags. In addition users should be able to not only add and search tags but also edit and remove them as they wish. Figure 8.4 shows the three basic components of a tagging system.

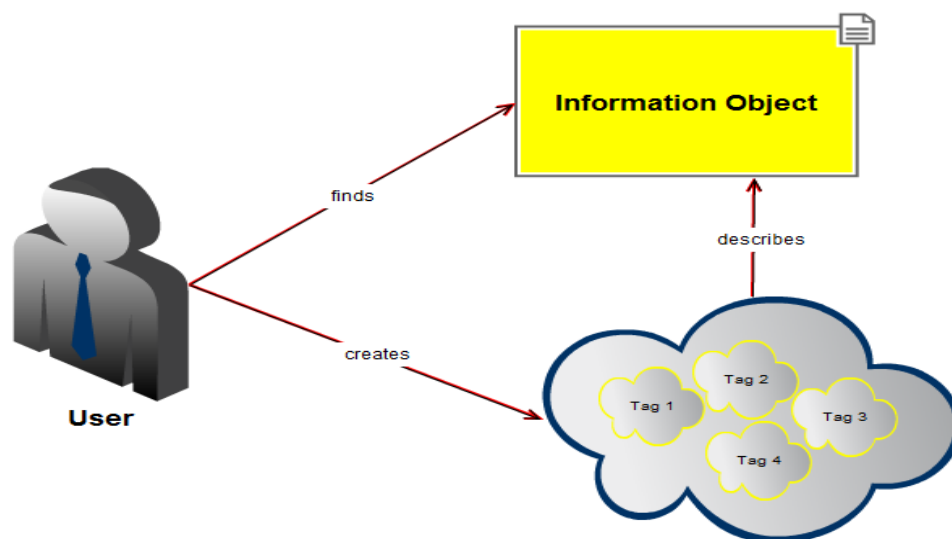


Figure 8.4: Components of a Tagging System

Interviewees raised some pressing issues associated with current tagging systems, including the lack of a structure, the question of synonym control and the absence of spelling controls. One interviewee (Study2-Interviewee4) recommended the use of term suggestions from existing tags as well as mapping user tags to a more formalised set of terminologies, as a solution. Although, it raises the question of how the first sets of terms are allocated, creating the “chicken-egg” dilemma.

Since tags are added by users themselves, the likelihood to match their information search keywords is much higher than if standardised and formal taxonomies are used (Weinberger, personal communication, December 12, 2012). Tagging in general places no constraints on

the use of terminologies, whereas taxonomies do (Study2-Interviewee14). It is difficult to anticipate the keyword users might use to describe a resource, or what keywords they will use to search. Whereas, taxonomies are mostly added from a controlled list wherein the values are mostly predetermined, and hence can be anticipated. But the flexibility of tagging, as two interviewees (Study2-Interviewee1; Study2-Interviewee18) identified, may result in poor and irrelevant metadata, which they believed should be cleaned up by librarians, given that resources are available to rectify such issues. They also mentioned the problems of tags that are too personal, such as ('awesome', 'to be read', 'mum's gift', and 'nice'). Some interviewees suggested the use of controlled vocabularies, such as subject headings, so that the nearest values from user tags can be mapped to more formalised and relevant tags. However, this creates an important challenge, regarding what is to be meant by 'correct' metadata (for more on metadata quality control see 8.6, and for metadata challenges see 8.7). Since users mostly provide keywords that represent information objects, tagging in general can thus be subsumed under the descriptive metadata category, rather than, for example, under technical or structural metadata. That means when users add tags, they are providing their own interpretation of the information object.

Thus, from the Open Coding interview data analysis, tagging emerged as an important instantiation of the socially-constructed metadata platforms (see also Figure 8.5).

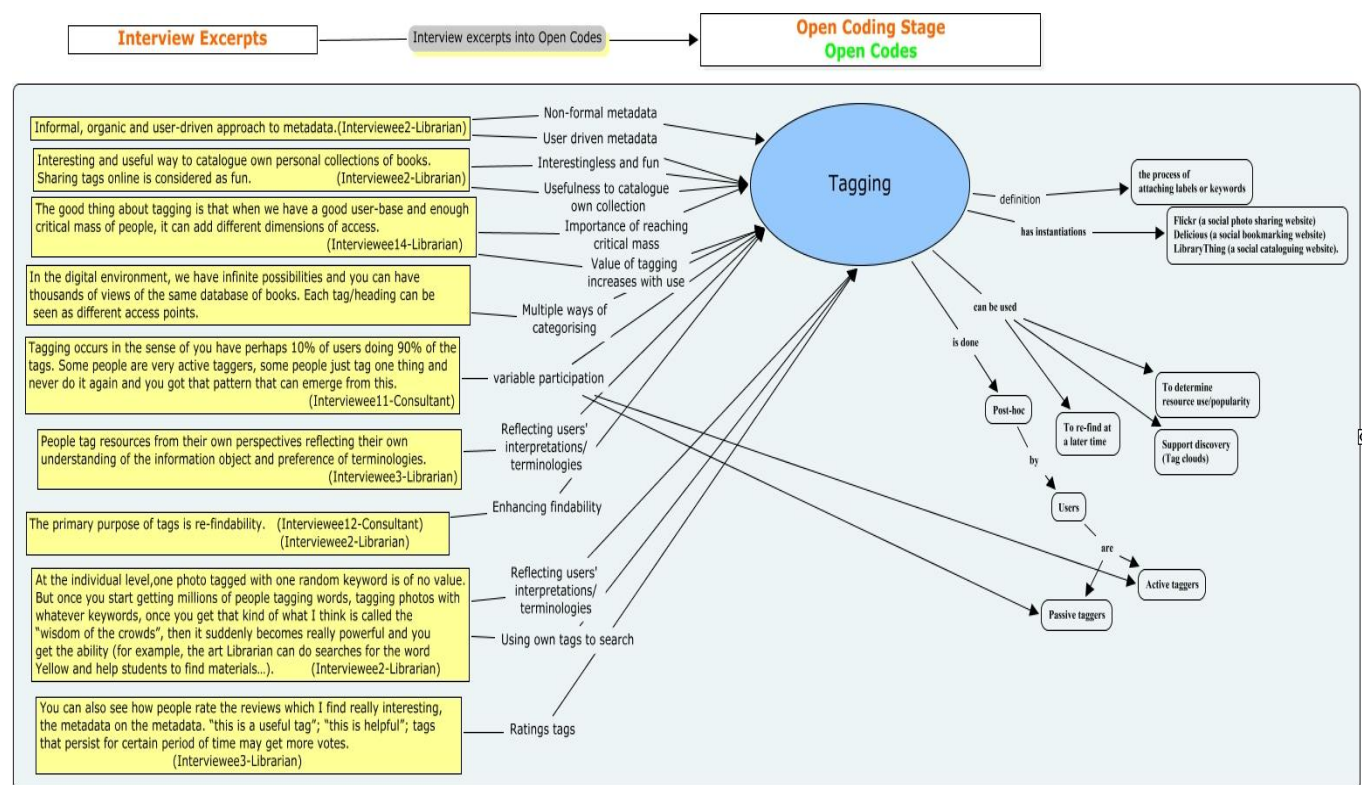


Figure 8.5: Tagging (Open Code during Open Coding)

8.4.1.2. User Reviews

Reviews are user submitted comments and critiques, concerning information objects. In relation to user reviews, one interviewee (Study2-Interviewee2) mentioned how they conceived of the idea of rating user reviews themselves (metadata of metadata) as particularly interesting. User reviews on Amazon, they pointed out, are very helpful for librarians who used to spend significant amount of time looking for book reviews on journal and newspaper archives, which are too formal to be helpful and also have very few books covered in such traditional media. The interviewee argued that reviews could be used in a diverse number of ways in libraries, when attempting to improve the services.

Most interviewees concurred with the importance of user reviews and their potential role in library metadata. However, the subjective nature of reviews was noted, as “it carries certain value judgement, certain value criteria which you may choose to trust or not” (Study2-Interviewee12). The relevance of the review is usually dependent on who writes the review. One interviewee noted the usefulness of provenance metadata as a means of identifying who has written it, saying for example “this review of a physics book is submitted by a third year under graduate in art history” (Study2-Interviewee6). As libraries collect a large variety and diversity of information objects, the importance of socially-constructed metadata by way of reviews is crucial. Such metadata helps users to identify and determine the relevance of information objects. Review metadata can also be structured and merged with other forms of metadata.

Thus, the Open Coding of interview data analysis helped to identify ‘User Reviews’ as one of the Open Codes (Figure 8.6).

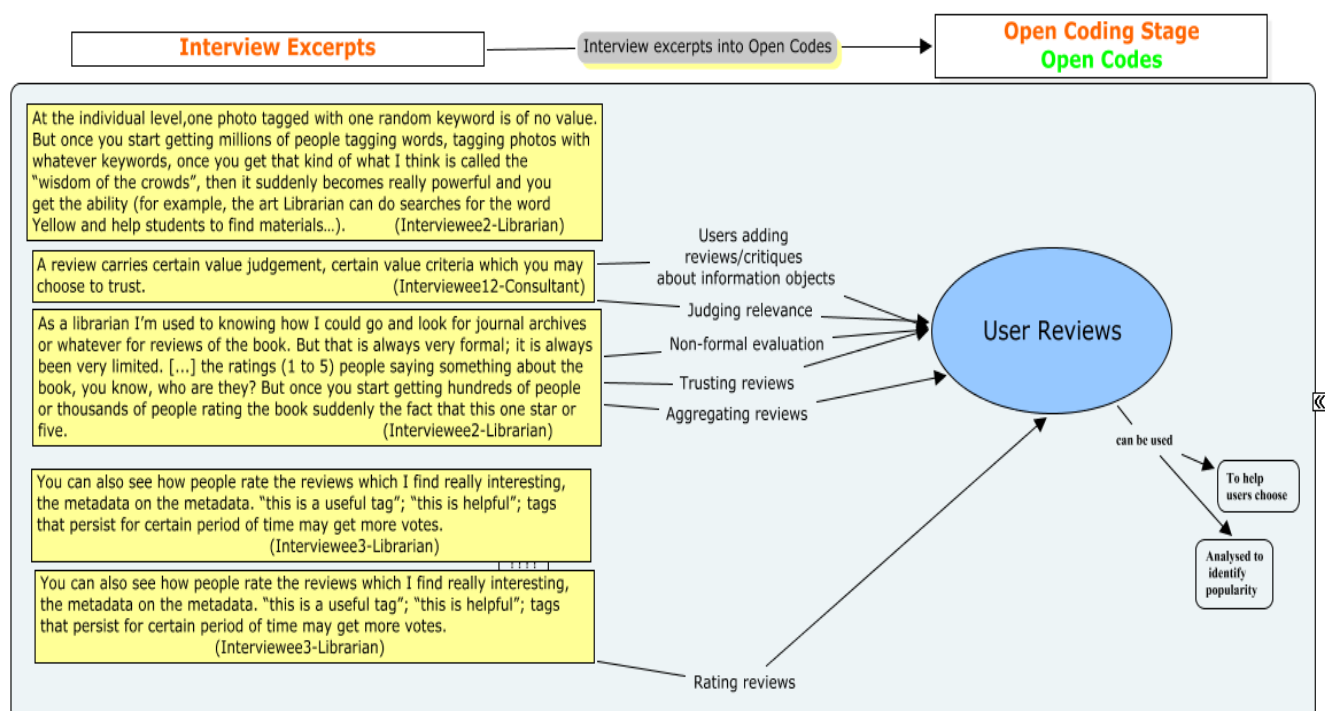


Figure 8.6: User Reviews (Open Code during Open Coding)

8.4.1.3. Ratings

Rating is an automatic means of determining users' preferences for information objects. Ratings can provide useful information regarding popularity or usage patterns. The interviewees identified user rating as one of the techniques used for collecting data about user's views on a particular information object (see also Figure 8.7).

Rating of reviews and tags is considered by interviewees as 'metadata about the metadata', which the interviewees thought could also help in determining the veracity and quality of socially-constructed metadata. The value of ratings, as interviewees indicated, should be considered in combination with other types of metadata. Ratings may indicate popularity, but can as well be used as a way of social filtering, where values are aggregated and analysed thus providing relevant insights to determining the use of particular information objects. Rating can also be used as a facet to sift through large collections of information objects.

Thus, the Open Coding stage of interview data analysis resulted in the identification of the 'User Ratings' as an **Open Code**, which is considered as one instantiation of a metadata platform that supports user metadata co-creation.

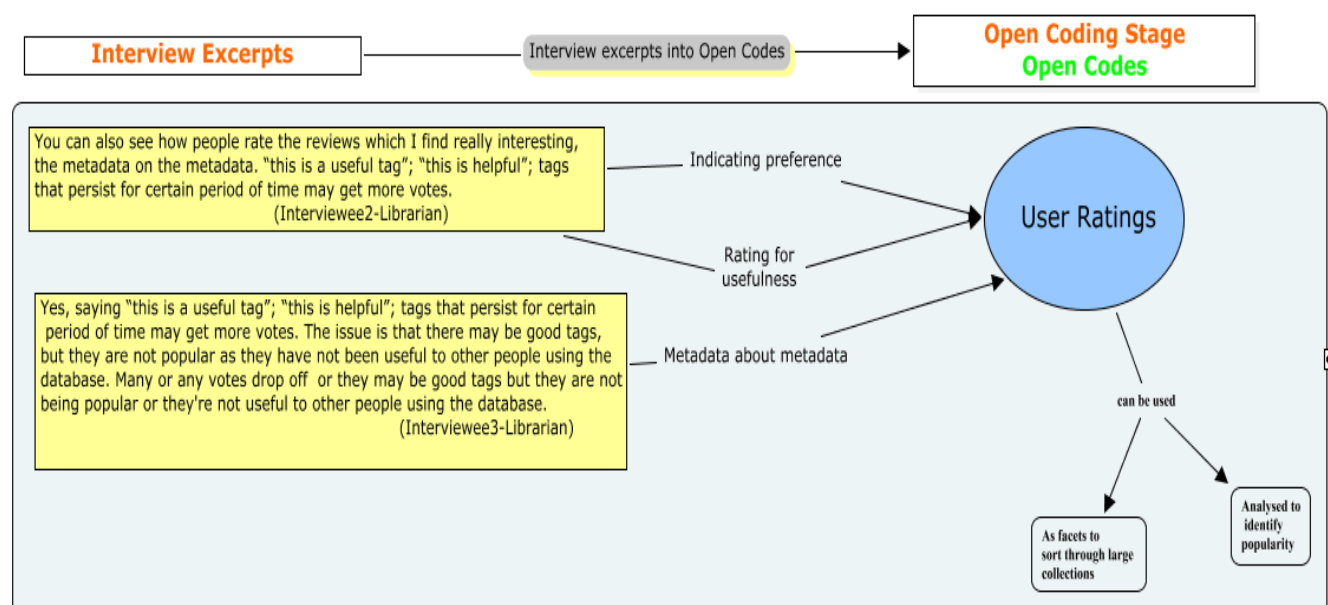


Figure 8.7: Ratings (Open Code during Open Coding)

8.4.1.4. Recommendation Systems

Recommendation systems are manual or electronic means for suggesting or sharing an information resource, on the understanding that the shared resource can be of some interest to the person who receives the recommendation. As regards to recommendation systems, one interviewee (Study2-Interviewee2) mentioned Amazon's 'the person who bought this one also bought this one' feature, which they said could be implemented in libraries as well. Most

other interviewees ascribed to the importance of recommendations from their peers and colleagues, which they believed is also instrumental in their online experiences. Thus indicating the potential of implementing, as one interviewee (Study2-Interviewee2) suggested, features such as “These students on this course have borrowed these books”. One interviewee (Study2-Interviewee10) emphasised the importance of people-to-people recommendations as networks of interest, which could be emulated digitally (see also Figure 8.8). Some of these recommendations, the interviewee indicated, require understanding the interests of others to whom the recommendation is catered for, thus tailored but nonetheless serendipitous services could be designed.

Whilst, most interviewees agreed with the importance of recommendations, some suggested that libraries should maintain the anonymity and privacy of users’ personal profiles, for example one interviewee foreseeing “situations where a particular user may decide their profile is going to be a political one” (Study2-Interviewee9), thus pointing to the crucial importance of underpinning such user profile, circulation or browsing data use policies on proper consultation and transparency with users.

Thus, during the Open Coding stage of interview data analysis, the Open Code ‘Recommendation Systems’ was identified (see Figure 8.8).

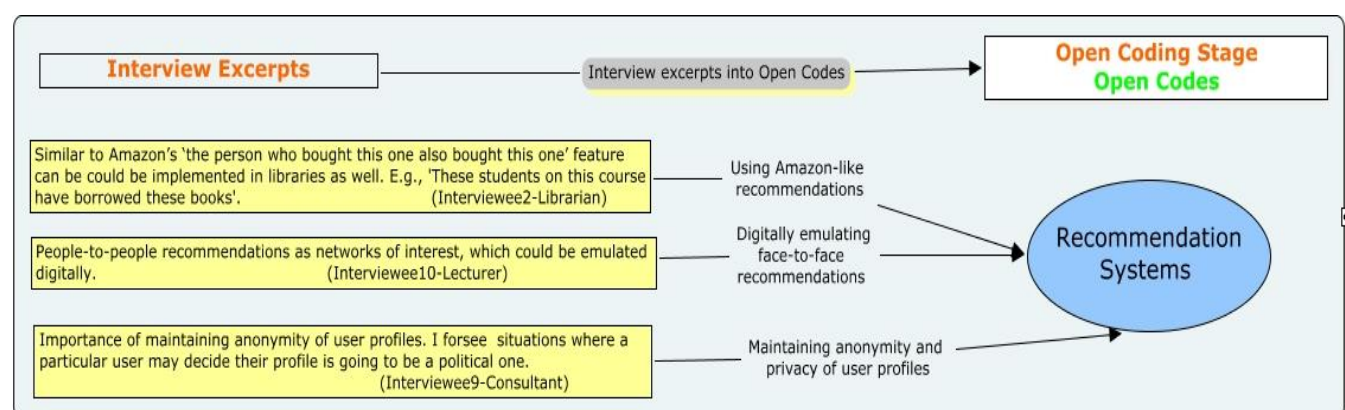


Figure 8.8: Recommendation Systems (Open Code during Open Coding)

8.4.1.5. Metadata Crowd Sourcing

Crowd sourcing is the process of announcing and soliciting volunteers to participate in specific projects and is considered as one of the core implementations of the Web 2.0 paradigm. Most interviewees cited the National Library of Australia’s Newspaper Digitalisation project as a good example of crowd sourcing, in which members of the public were asked to identify and correct digitised text which the Optical Character Recognition (OCR) technology could not recognise.

Other examples mentioned by interviewees include the Old Weather project, to add metrological data from old naval logs and Galaxy Zoo, a volunteer social scientific project, where anyone with an interest in astronomy contributes by identifying, classifying and discovering galaxies. One interviewee saw crowd sourcing as “a very powerful way to engage the public in what librarians do, getting people involved in a direct way, and allowing

people to contribute for the common good” (Study2-Interviewee10). So in this respect, metadata that is created by experts using standards-based metadata approaches can be further enhanced and improved. One interviewee stated that “we are at this point looking into options for a combination of crowd sourcing and metadata creation. We are definitely interested in the possibilities it might offer, and we want to do some pilot work with interested third parties on that just to see how far we can go” (Study2-Interviewee4). The interviewee had in mind the crowd sourcing of metadata creation for historical documents such as playbills. Experts in particular niche sources of information are scattered elsewhere and can be encouraged to describe particular documents at a high level of detail.

Interviewees indicated that libraries can collect metadata about information objects where the public has more knowledge than librarians. Some of the examples mentioned include identifying people in photographs. Metadata can thus be collected from the public about obscure information objects. One interviewee pointed out to a recent crowd sourcing project they have undertaken:

“We find in our digital library a historical picture but we could not make sense of it as no one among the librarians could identify who is the person in the photograph, and the time and location where the photos were taken. We then implemented a crowd sourcing project for a limited set of archival collections and we gathered very relevant metadata from people who were able to identify these pictures and its context in terms of places, names of people and other details. These were metadata which we could not have found otherwise. Sometimes you would be amazed how pertinent information can be gathered from users” (Study2-Interviewee21).

Thus, based on the Open Coding of interview data analysis, ‘**Crowd Sourcing**’ was identified as a potentially relevant metadata platform for user metadata co-creation (see also Figure 8.9).

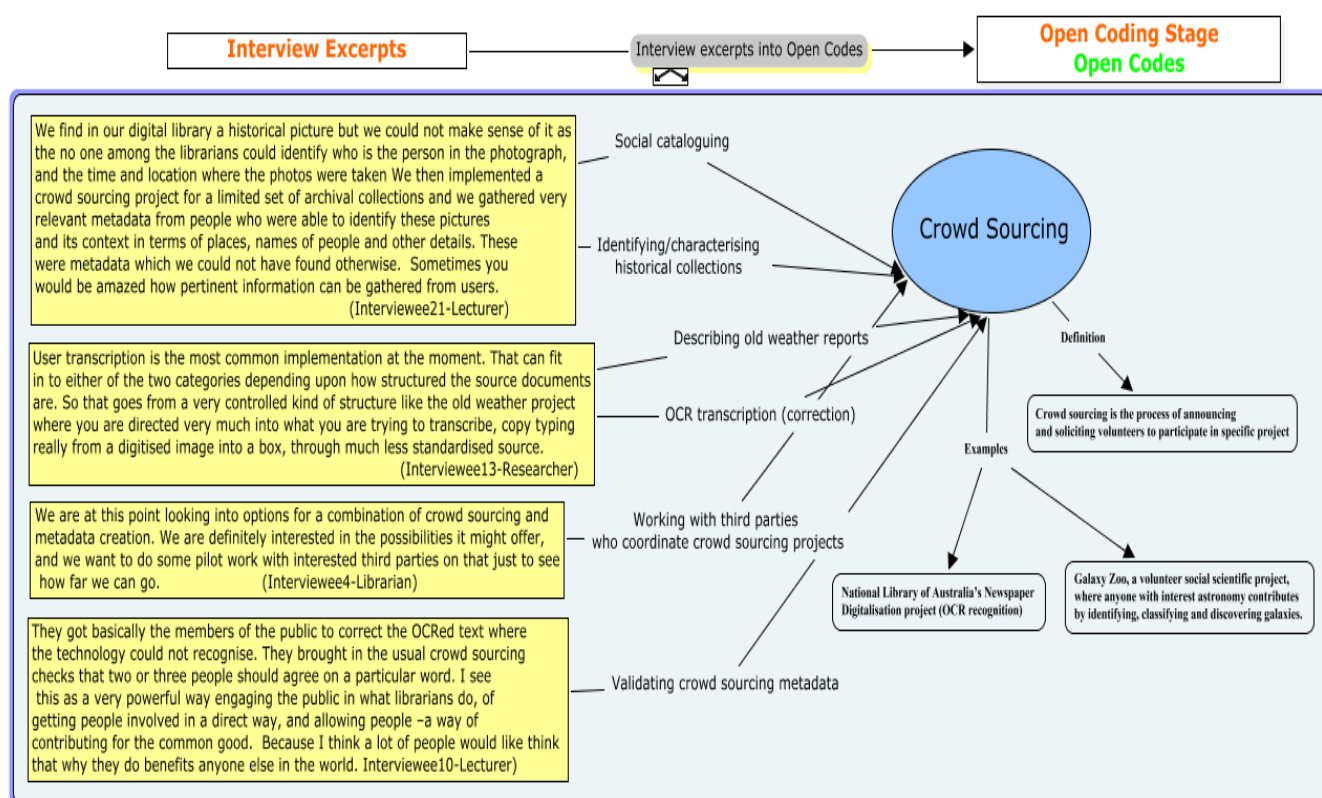


Figure 8.9: Crowd Sourcing (Open Code during Open Coding)

8.4.1.6. Relationship between the Open Codes of the Category of Platform for Proactive Metadata Co-Creation

As the data analysis showed, in order to enable users to contribute metadata, the importance of a technological platform that supports user participation is self-evident. As the responses from the interviewees indicated, this platform should not only allow users to consult and use content but should also enable them to proactively apply tags (structured labels), add reviews (unstructured critiques of information objects), rate (assessing relevancy of an information object) and recommend information objects to their peers (see Figure 8.10). This phenomenon was referred to by one interviewee (Study2-Interviewee8) as a new dimension to metadata, one which would bring users closer to libraries, through proactively engaging them in metadata functions. The same interviewee added that socially-constructed metadata approaches can be designed in such a way that users are involved in not only co-creating but also in controlling and managing metadata entries, with the result that it becomes a socially regulated and self-healing system (see 8.4.5.7). The interviewee believed that users are generally keen to get involved in altruistic and community-driven projects (see 8.5).

Such proactive involvement of users in adding metadata can be more aptly referred to as metadata co-creation, which signifies the ability of users not only to find, identify and retrieve information objects using metadata but also contributing new metadata descriptions, thereby adding new value to the system. Contemporary library management systems are primarily designed with the assumption that experts such as librarians create and manage

metadata and, hence, do not envisage user contributions. Thus, as one interviewee (Study2-Interviewee13) noted, “there was no space for putting user-contributed content within the description standards”. Interviewees therefore underlined the importance of revisiting contemporary approaches so as to retrofit them to present-day requirements, especially in view of the need for engaging users as co-creators.

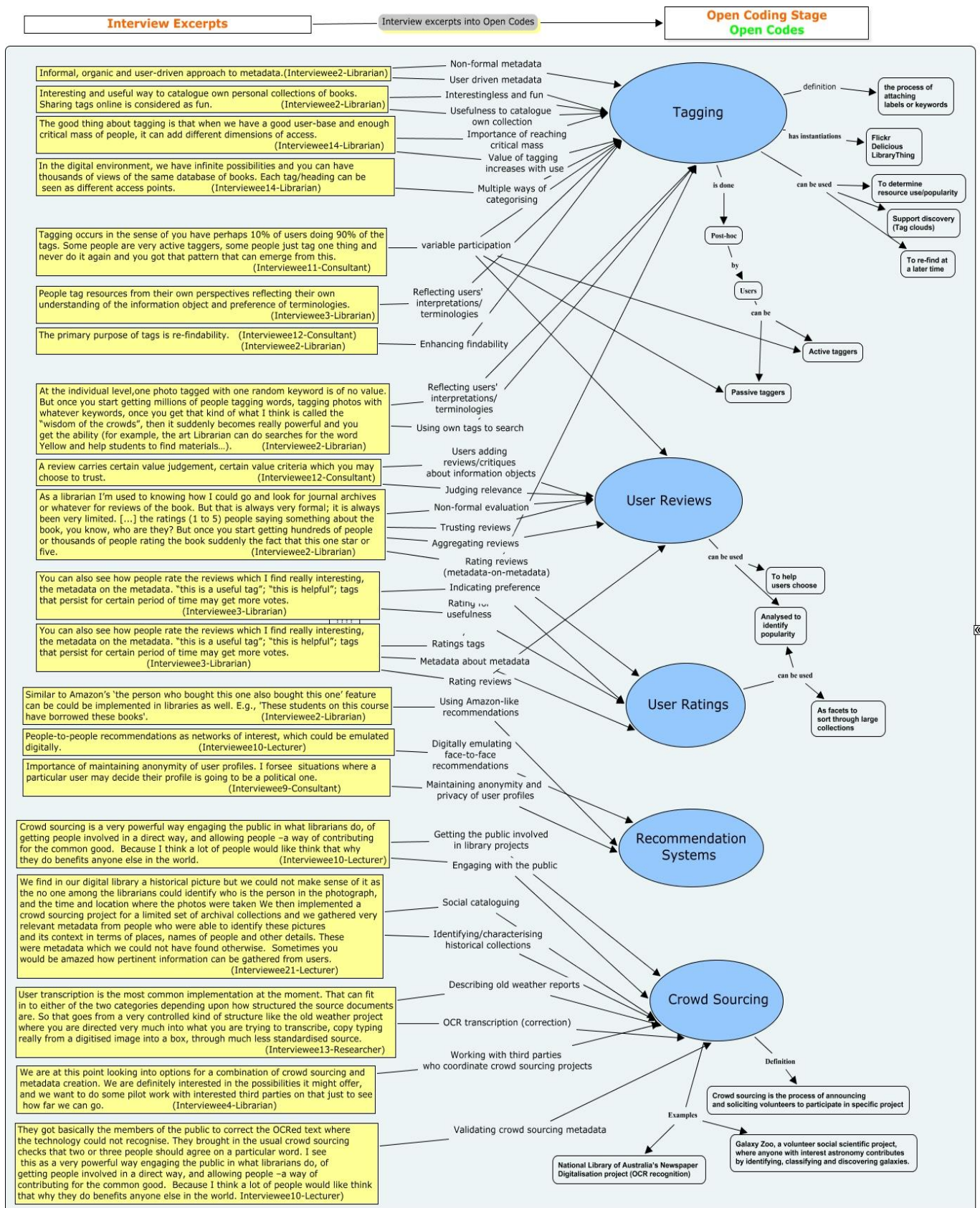


Figure 8.10: Open Codes relating to Platform for Proactive Metadata Co-creation

8.4.2. Users as Proactive Metadata Co-creators

As discussed in section 8.3, socially-constructed metadata approaches present an opportunity for libraries to involve users as metadata creators. Based on the narratives in 8.3 and also based on the Open Codes (see Figure 8.11), the Focused Coding of interview data analysis showed that unlike standards-based metadata approaches, in socially-constructed metadata approaches users can be considered as partners in the creation and utilisation of metadata.

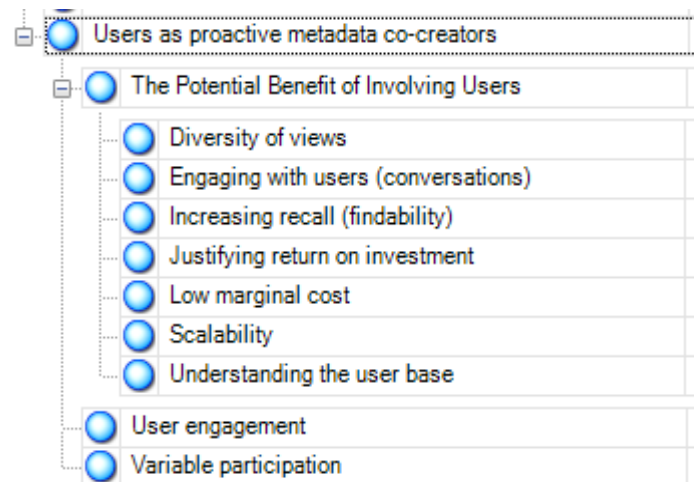


Figure 8. 11: Open Codes relating to Users as Proactive Metadata Co-creators

The use of socially-constructed metadata approaches provides libraries the opportunity to better engage with their users (Study2-Interviewee12). This is because the platform presented to users (8.4.1) would be an encouraging and inviting one for users to participate and engage. Thus, users can be proactive metadata creators. Once librarians in general, and cataloguers in particular, assumed the exclusive role of affixing metadata to books and related information objects. However, there now seems a recognition that librarians should relinquish their sole control of such functions. In connection with this, one researcher (Study2-Interviewee20) underlined the importance of treating users as partners and pro-active metadata co-creators rather than mere ‘consumers’ of information. Without such pro-active participation (Figure 8.12) of users, the interviewee warned that, libraries will likely lose their place as major information service providers.

“The problem with some librarians I know of is that they want to be the only ones. They were Fifty years ago. Then, when you want information, the only place you could go is to the library but this privilege does not exist anymore. But now, librarians do not have the exclusivity of providing information services. But this is not bad either. Because the librarian’s role has now improved, the role of the librarian is no more constrained to the physical library and its collection and organisation. I believe the role now is much better than it was when librarians had to deal with only the physical collection. I want to stress that, as LIS professionals, we should not expect exclusivity and we should not assume we are [the only ones] to offer information services to our users” (Study2-Interviewee20).

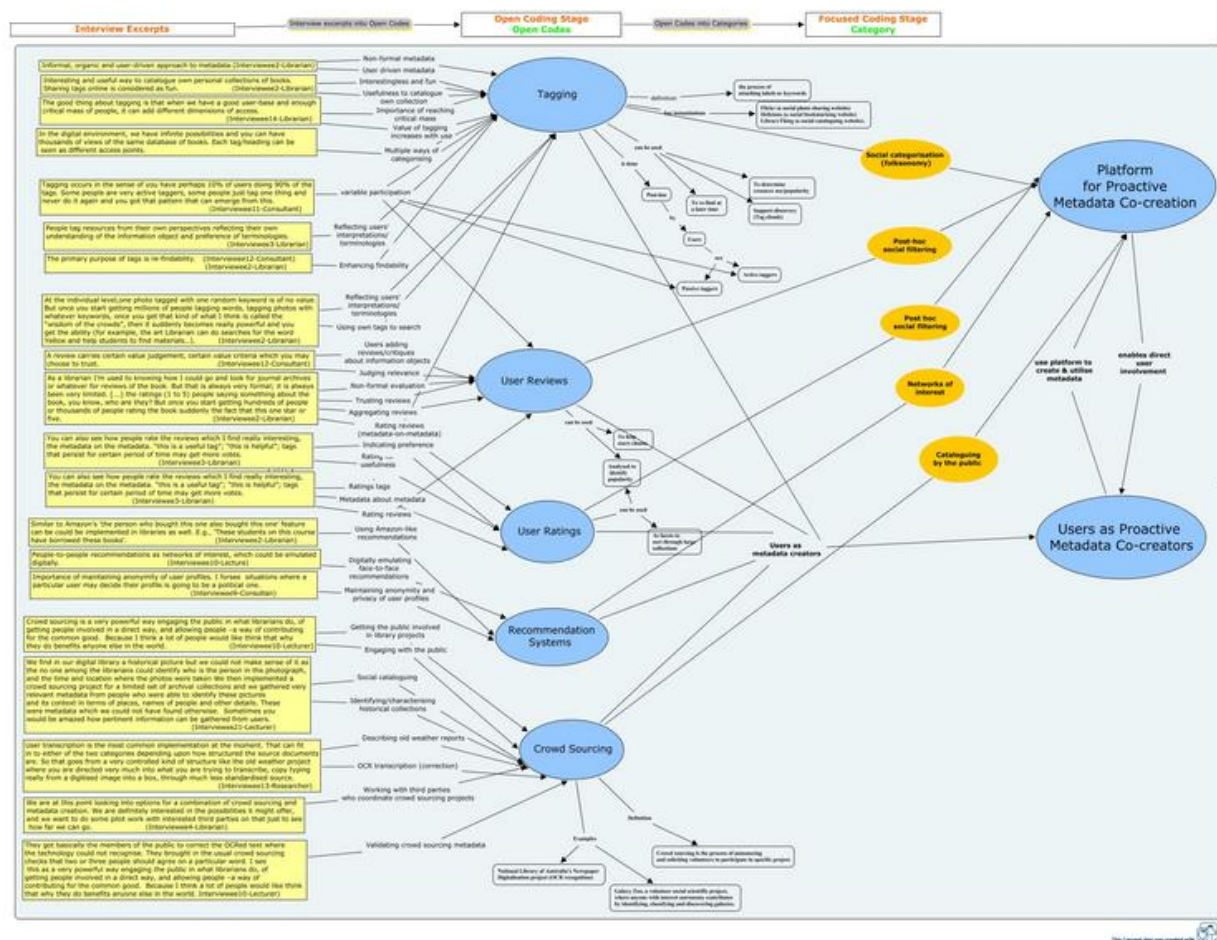


Figure 8.12: The Category of Users as Proactive Metadata Co-Creators

8.4.3. Metadata Diversity

The platform for proactive metadata co-creation (8.4.1) allows users to add metadata (8.4.2) thereby enabling the addition of multiple metadata entries about the same information object; each of which would bring in new descriptions, novel view points, and varying subjective preferences. For the metadata platform to permit diversity, it should allow a variable amount of participation. The platform should be able to aggregate (see 8.4.5) and make sense of the variable contributions (8.4.5.1), so that additional value for other users comes from this aggregation and not from the single entry. The interviewees identified the Long Tail (8.4.5.2) and Wisdom of the Crowds (8.4.5.6) as important concepts that need to be considered in light of metadata diversity. One of the benefits of metadata diversity is the representation of not only similar but also contrary interpretations of information objects.

One of the categories that emerged from the analysis is “emergent vocabularies of users” which interviewees consider as an important component of metadata. All interviewees were cognisant of possible discrepancies between the scientific terminologies used in standards-based metadata approaches versus the day-to-day use of terms by users for searching databases. To overcome this, as interviewees mentioned, many users seek assistance from librarians for guidance in finding resources from library catalogues. Interviewees saw no

reason of why the terminologies of experts and users should not be combined and used, thereby enhancing the findability and discoverability of information objects.

The interviewees also recognised that metadata should reflect the emergence of new vocabularies, adding that users can contribute greatly in bringing this about. In addition, metadata diversity should also consider sociological differences, so that each group's and individual's use of terminologies is represented as adequately as possible. One interviewee (Study2-Interviewee8) pointed out the importance of collecting both academic/scientific metadata from a community of scholars and "unstructured" metadata from regular library users and readers. This interviewee in particular emphasised the relevance of incorporating people's opinions about films. According to the same interviewee, "there is no superfluous and irrelevant metadata; it is all dependent on context" (Study2-Interviewee8). Most interviewees mentioned Wikipedia as an example where there are diverse groups of contributors, both professionals and, non-academic or lay contributors. Citing their own experience, many interviewees noted that they see potential that experts will contribute out of professional duty. Thus, the Focused Coding stage of data analysis brings the Open Codes "The Long Tail of Metadata" and "Wisdom of the Crowds" into the Category of "Metadata Diversity" (see Figure 8.13).

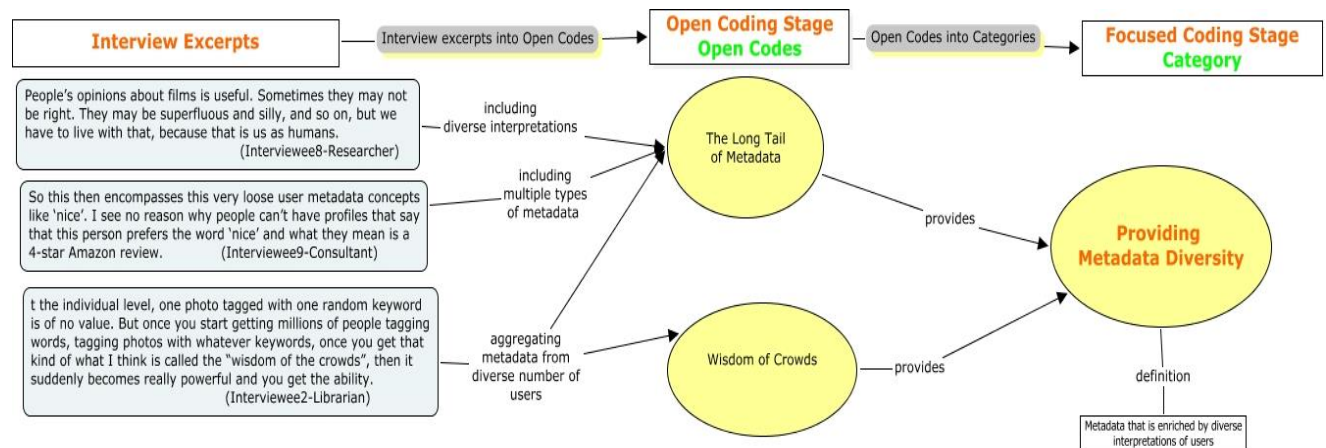


Figure 8.13: The Category of Metadata Diversity

8.4.4. Metadata Scalability

One of the challenges of standards-based metadata approaches, as noted by interviewees, was lack of scalability and inability to adequately describe the plethora of information objects found in libraries. Whilst standards-based (expert-created) metadata may not be able to scale to enrich information objects with diverse interpretations of users, socially-constructed metadata approaches might be able to meet the metadata need to adequately describe increasing library collections. One interviewee (Study2-Interviewee7) indicated that metadata scalability can be achieved through involving users. Similarly, another interviewee (Study2-Interviewee2) also stated that involving hundreds and thousands of library users is a worthwhile endeavour to address the metadata scalability issues libraries face.

Thus, during Focused Coding the Open Codes such as “Involving Large Number of Users” and “Diversity of Views” resulted in the identification of a Category called “Metadata Scalability” (see Figure 8.14).

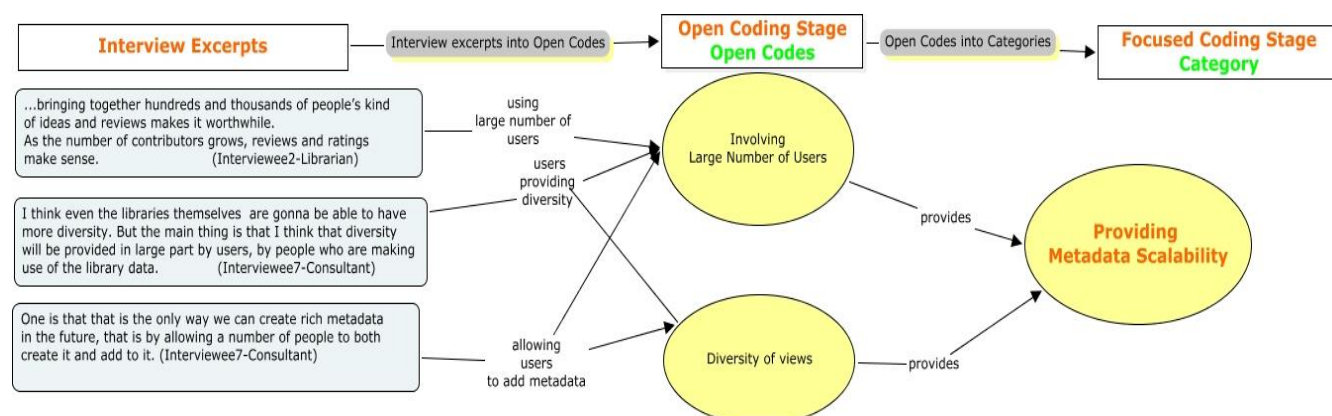


Figure 8.14: The Category of Metadata Scalability

8.4.5. Collective Metadata Intelligence

The Focused Coding analysis of interview data indicated that collective metadata intelligence is an emerging concept as users actively co-create metadata using a metadata platform which accepts variable contributions from diverse groups of users (see Figure 8.15). Collective metadata intelligence indicates that metadata entries about information objects are collected and aggregated, providing useful insight and patterns. The intelligence mentioned here refers to added value that can be gleaned from aggregating, mashing-up and analysing metadata. As the data analysis through Focused Coding showed, collective metadata intelligence includes concepts such as variable metadata participation, the Long Tail of metadata, metadata aggregation, network effect and the “Wisdom of Crowds”.

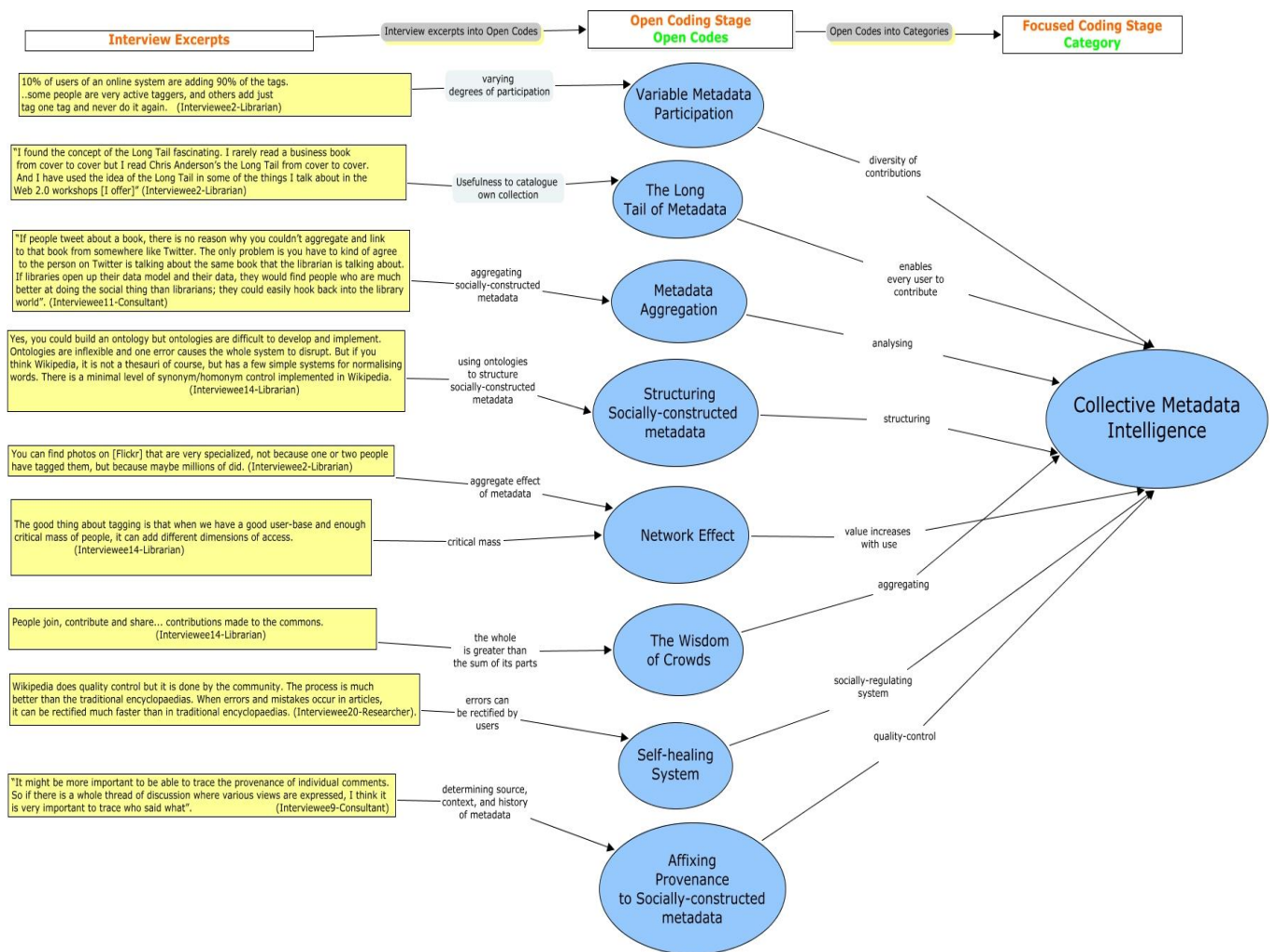


Figure 8.15: The Category of Collective Metadata Intelligence

8.4.5.1. Variable Metadata Participation

The interviewees believed that not all library users gravitate towards adding metadata but emphasised that even the contributions from the small number of users may in aggregate result in significant amount of metadata. The interviewees elaborated on the possible variations in metadata contributions by users. One interviewee (Study2-Interviewee4) alluded to the possibility of having only a small percentage of users taking an interest in adding metadata, whilst the great majority remained simply as passive users.

Interviewees mentioned the case of Wikipedia as a good example where there are only a small percentage of contributors in relation to the large number of users. There is thus variable participation (a phrase coined by Shirky, 2008), where some users are very active and prolific contributors, whilst many others may simply create a single article entry and others are in between. In this regard, one interviewee (Study2-Interviewee4) estimated that only about 10% of users of an online system are adding 90% of the tags. The interviewee further commented that, “some people are very active taggers, [but others] add just one tag and never do it again”, thus indicating variable participation. But given the possible high

number of users in a given information system, even the few number or active taggers would constitute a large set of tags.

8.4.5.2. The Long Tail of Metadata

Attributing the concept to Chris Anderson (Anderson, 2006), interviewees identified the Long Tail as an important emerging approach to metadata. One interviewee stated:

“I found the concept of the Long Tail fascinating. I rarely read a business book from cover to cover but I read Chris Anderson’s the Long Tail from cover to cover. And I have used the idea of the Long Tail in some of the things I talk about in the Web 2.0 workshops [I offer]” (Study2-Interviewee2).

The interviewee also recalled a talk which they had attended at a conference where the speaker recited Amazon’s use of the idea of the Long Tail where the company reportedly derived its estimated sale of 55% out of niche (obscure) items rather than blockbusters and best sellers. The concept of the Long Tail, the interviewee noted, could be contextualised to libraries both in their collections of information objects and also associated metadata description. Reaching the long tail of collections refers to including in the library’s collection not only heavily used items but also those that may have only a few users in which the individual needs of users can be optimised. For metadata this means, including any tags, reviews, and ratings from all contributors without focusing on tag clouds (highly tagged or highly searched items), thus each contributions can aggregate, thus increasing the value of metadata.

8.4.5.3. Metadata Aggregation: Tag Clouds, Tag Clusters and Tag Patterns

As users tag information objects with their chosen terminologies, each term remains flat (unstructured without synonym or homonym relationships) and uncontrolled. Individual tags offer meaning to the person who contributes it rather than to other users. However, in order for tagging data to make sense to other users, the tags should be aggregated and analysed so that tag clouds and tag clusters are generated. The social value of tagging comes therefore from its aggregation. There are various ways of aggregating the tags (see Figure 8.16). Metadata aggregation implies the value of what is referred to by the interviewees as the network effect of metadata (8.4.5.5).

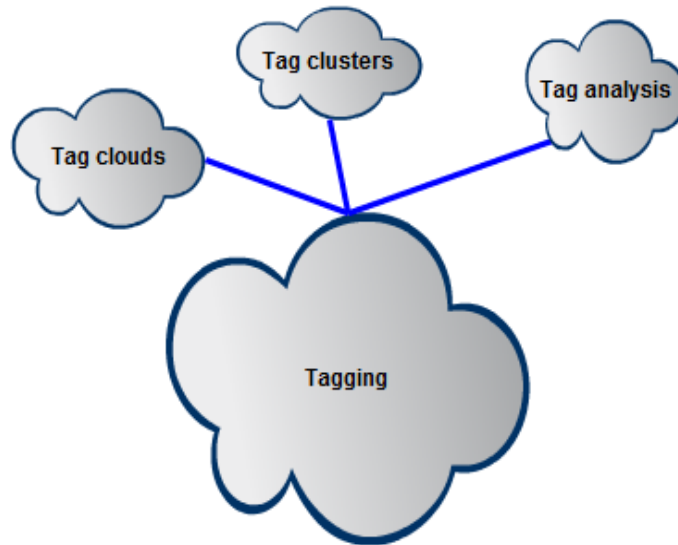


Figure 8.16: Metadata Aggregation

Tag-clouds are considered as an interesting development in social tagging where the system generates the frequently appearing words and phrases in a document. Tag clouds are represented as images of the words where words looking bigger indicate they have appeared more frequently than the ones shown in smaller sizes. As one interviewee discussed tag clouds could be generated for any electronic document, such as a blog entry. A tag cloud can provide a hint at what an information object is about. Most interviewees were not particularly keen in using tag clouds. One interviewee stated that “the way I would tend to use a tag cloud is if I am into a completely new site” (Study2-Interviewee12). Another interviewee said “I’ve used them in the past. It is quite easy to get lost in a tag cloud as you click through them. Before you know it you are off into some little avenue somewhere” (Study2-Interviewee6).

The interviewees indicated the importance of using contextual tag clustering so as to make sense of the flat tags added by users. A kind of tag cross-mapping between concepts, terms and representations is suggested, for instance as one interviewee (Study2-Interviewee9) suggested there may be a relationship between what it means when a user rates an information object as “5 stars” in one application such as Amazon to rating it as “very good” in another application. It is also suggested that clustering tags based on context and semantics is important. Another interviewee (Study2-Interviewee3) suggested the use of subject headings and thesauri and mapping tags into an existing taxonomy (classification).

Metadata aggregation can be made not only within a particular information system or application but also across applications and services. For instance one interviewee believed that:

“If people tweet about a book, there is no reason why you couldn’t aggregate and link to that book from somewhere like Twitter. The only problem is you have to kind of agree to that person on Twitter is talking about the same book that the librarian is talking about. If libraries open up their data model and their data, they would find

people who are much better at doing the social thing than librarians; they could easily hook back into the library world” (Study2-Interviewee11).

Metadata that comes from users through ratings can be analysed providing meaningful results into the preferences and zeitgeist (such as popularity of some information objects over others). Users leave significant amount of implicit metadata as they use information systems such as online databases. The clickstreams provide an important insight into their use behaviours and frequently viewed information objects. Whereas such metadata is important, in order to glean a meaningful result, the clickstreams of as much users as possible should be collated and analysed. The value proposition of socially-constructed metadata approaches is thus heavily hinged upon the aggregated effect from which individual contributions within and across applications can be aggregated, analysed, structured and collated in a meaningful way, thus adding value to the experience of users.

8.4.5.4. Structuring Socially-Constructed Metadata

Socially-constructed metadata, in the form of comments and reviews comes in an unstructured format, whereby it cannot be directly made to correspond with any structured schema of fields and constraints. It is therefore important that technological means be explored for mapping unstructured information into structured data that is amenable to machine-processing. As users rate and vote for information objects, the rating data should be aggregated, and its mean analysed so as to help indicate the most popular information objects. One of the interviewees (Study2-Interviewee16) noted the importance of using the structure of existing standards-based metadata to make sense of socially-constructed metadata.

8.4.5.5. Network Effect

Network effect is the value and utility of socially-constructed metadata increases as the number of contributors and hence the size of metadata also increases. As the interviewees pointed out, the utility and value of specialised and personalised tags only becomes apparent as the numbers of contributors increase in such networks. As they indicated, a network effect of metadata has been demonstrated with online social applications such as LibraryThing and Flickr. As the number of contributors grows, more metadata is created and patterns on the preference and use of terminologies, popular usages, and zeitgeist could begin to emerge. Whilst individual metadata values of socially-constructed metadata may not necessarily provide useful insights, the network effect of metadata however plays an important role. Put another way, metadata value increases as more metadata exists (Study2-Interviewee2; Study2-Interviewee3; & Study2-Interviewee9). The interviewees in particular mentioned the law of large numbers or the “Wisdom of Crowds” as important techniques to glean benefits from the network effect. One interviewee (Study2-Interviewee2) in particular exemplified the network effect by saying, “you can find photos on [Flickr] that are very specialised, not because one or two people have tagged them, but because maybe millions did”.

Similarly, others also cited Wikipedia as an example of the network effect, where each additional article, each edition and improvement add value to it. Singling out each contribution, may not make the value so apparent, thus, as Study2-Interviewee2, advised, the

aggregate effect should be looked at. The interviewee noted their experience thus, “I search for really odd things, fairly sure that I’ll find something because there’ll be some insane person out there that’s tagged their photo with shoe-lace, lost in the high street, and something really off the wall. You’d think ‘no one would have ever done that under the formal system’, but tagging kind of allows that kind of metadata to suddenly be free, and of course computing power has come along allowing us to actually process that”. Similarly, another interviewee (Study2-Interviewee14) added “the good thing about tagging is that when we have a good user-base and enough critical mass of people, it can add different dimensions of access. We librarians know that the idea of DDC was born out of the need to place one book in one place, one book at one shelf, because that is how atoms work. In the digital environment, we have infinite possibilities and you can have thousands of views of the same database of books” (Study2-Interviewee14).

8.4.5.6. The Wisdom of Crowds

When variable individual user metadata contributions (8.4.5.1) are aggregated, it is indicated, the value of the collective metadata increases. One interviewee (Study2-Interviewee3) attributes this to the “Law of Large Numbers”, whilst many others referred to it as the “Wisdom of Crowds”. As one interviewee (Study2-Interviewee2) noted the contributions make sense as more people continue to add metadata. Interviewees recited the examples of Wikipedia, Amazon and Flickr as good examples of the “Wisdom of Crowds”. The variable user contributions become structured, aggregated, analysed, and collated, resulting in what the interviewees alluded to the “Wisdom of Crowds” (a concept originally attributed to Surowiecki, 2004). As one interviewee (Study2-Interviewee14) noted, one of the motivations for users to contribute socially-constructed metadata is the recognition that their contributions “collectively benefit everyone”.

One interviewee (Study2-Interviewee2) pointed out that the “Wisdom of Crowds” was a powerful means for collecting metadata from users. Adding metadata descriptions to the ever growing collections of digital documents can only be accomplished by involving an increasing number of users, thereby overcoming resource and time constraints.

The notion of the “Wisdom of Crowds” is thus an important element of socially-constructed metadata approaches where, “people join, contribute and share” (Study2-Interviewee14), thus everyone would benefit from contributions made to the “commons”. The “Wisdom of Crowds” refers to the insight, understanding, richness and knowledge that may be derived from the aggregate effect of metadata that has been collected from the Long Tail (8.4.5.2) through the network effect (where value of metadata system increases with new users joining the system and contribute metadata).

8.4.5.7. Self-Healing System

In view of improving the quality of socially-constructed metadata, interviewees noted the potential challenges it presents to institutions to vet, police, edit or improve each metadata. Interviewees suggested that a self-healing system where users themselves should be involved in partnership with metadata experts to ensure the smooth functioning of such systems,

including reporting malicious entries and improving obvious errors. Most interviewees cited Wikipedia as a good example of a social project where it has mechanisms in place to regulate its content. Two interviewees indicated that, due to the enormity of size of articles on Wikipedia, a single organisation could not possibly manage the quality of the content; instead Wikipedia introduced a self-healing system where users themselves participate.

One interviewee remarked that, “Wikipedia does quality control but it is done by the community. The process is much better than the traditional encyclopaedias. When errors and mistakes occur in articles, it can be rectified much faster than in traditional encyclopaedias” (Study2-Interviewee20). Similarly, the interviewee suggested a similar approach in academic libraries where mechanisms are found for users to partner with librarians to contribute in the management of socially-constructed metadata. The interviewee further commented that librarians are not always experts in subject domains, thus suggesting that there is much for libraries to benefit by involving users in the improvement of metadata generated by users. Partnership with users as peers rather than distant consumers of information, the interviewee noted, can provide libraries with a competitive advantage over other information providers. Most interviewees recognise that user involvement in metadata management implies that librarians should give up some of the traditional control on metadata.

8.4.5.8. Affixing Provenance to Metadata

The concept of provenance which was considered as one of the categories of standards-based approaches (7.4.3) was also considered an important component. Metadata provenance, which includes information about ownership, authority, history, and context of metadata, provides users with a set of additional information about specific metadata statements (assertions) in relation to a given information object. Whether it is a review, tag or recommendation, provenance plays a significant role in addressing some of the metadata quality concerns of librarians. One interviewee indicated that, “once you add provenance data; I don’t think librarians should be in any sense, assessing the quality of it and to say this is good or this is bad, it is up to other users to determine whether a metadata statement is good or bad. But they have to know who said it, so back to provenance again” (Study2-Interviewee9). One interviewee (Study2-Interviewee13) noted: “It might be more important to be able to trace the provenance of individual comments. So if there is a whole thread of discussion where various views are expressed, I think it is very important to trace who said what”.

Provenance was mostly discussed in relation to standards-based metadata. However, the interviewees considered it as an important component which can provide context to socially-constructed metadata. Provenance, as the interviewees suggested, provides transparency to metadata. Some interviewees considered it as metadata about metadata. Provenance is considered thus as one of the mechanisms to show the authenticity and credibility of metadata, wherein users are kept informed with regard to the source and context of metadata.

8.5. Motivation for Socially-Constructed Metadata Approaches

Metadata creation and maintenance in libraries has been considered the duty of metadata experts. If socially-constructed metadata approaches are to be implemented, what motivations exist that persuade users to describe information objects? The motivations that drive users to contribute metadata vary widely, depending on the type of application and the intended purposes of the metadata. For example, as one interviewee recalled, their library successfully accomplished a small project that involved describing old maps; they found many enthusiastic people who were willing to describe the maps with detailed metadata, which otherwise would have been difficult for in-house cataloguers to add such rich metadata by themselves in such a short time. The interviewee (Study2-Interviewee5) admitted that it was very difficult to accurately determine the exact motivations of users, but was of the opinion that such motivations may include personal professional interest, sense of belongingness to the institution and altruism.

Another interviewee discussed their own experience that people are willing to contribute, depending on the topics and disciplines and also the goals of the task requested:

“If people are interested in a domain, people become very supportive of activities in that particular area; for example some people may be very knowledgeable about laptop computers and know them inside and out. Or they know about a particular genre of orchestra and know it inside out. And when they come across a particular inaccurate detail about a particular laptop or an orchestra, they feel the urge to correct and elaborate. So we just have to let things take their natural course. Most people have the altruistic awareness that they operate in a community. Self-balancing will then occur” (Study2-Interviewee8).

Another interviewee (Study2-Interviewee10) indicated that they saw no great problem in finding people with enough motivations to contribute to socially-constructed metadata. The interviewee stated that “cataloguers should still be producing very high quality cataloguing data, but I think users, educated users, people who worked for the social good in their own careers, they are all going to be motivated to supply good quality metadata. We just have to give them the infrastructure and the tools to do it”. The interviewees mentioned the example of Wikipedia, where people, out of various motivations, do contribute articles and also volunteer to maintain its quality. Lowering barriers to entry, simplicity of interface and interestingness of purpose, re-findability, sense of ownership, altruism and engaging users were considered by interviewees as the major drivers of motivation for users in participating in socially-constructed metadata approaches.

8.5.1. Reducing Barriers to Contribution

Users are more likely to contribute to a metadata system that has low barriers to entry. This refers to technical complexity. This is in relation to the interface of the system and also the requirement to login. One interviewee stated that:

“One needs to be careful about [asking the user too much by way of login details] as it might be a barrier to participation in the first place. You have to collect the minimal amount of background information about users that is appropriate for the project you have in mind” (Study2-Interviewee13).

The interviewees indicated that, Wikipedia, for example, is said to have a low barrier to entry, whereas contributions to a peer-reviewed journal face a considerably high barrier to entry, mainly due to the review and editorial controls put in place. Most interviewees suggested that socially-constructed metadata systems should be flexible and open enough to allow novice contributors. In other words, they seemed to suggest that, systems should not enforce a very stringent set of data encoding (metadata entry) requirements. It is indicated that some applications have a rigorous user registration, pre-filtering and approval (confirmation) process that users may not necessarily want to undergo, unless the return on investment of their time spent passing all these procedures is high enough. Interviewees thus proposed that metadata systems should have low barriers to entry for users. That is, the system should be usable and content should be interesting. In addition, the interview data suggested that, the learning curve for metadata system interfaces should also be very low. Put another way, users should not be made to spend lots of time learning how to add metadata into the system, as they may expect it to be seamlessly interwoven with the OPAC/discovery system. Hence, they should not be asked to login in a separate interface.

In order to reduce the barriers to entry, the stringent metadata controls that are put in place in standards-based metadata approaches should be reconsidered. The implication is that metadata experts may be required to relinquish some of their power in metadata control and share some of the responsibility with users, so that users feel both empowered and get a sense of ownership (8.5.4) for the metadata that they create.

8.5.2. Simplicity, Interestingness and Fun

In order for users to add metadata, the social application platform should appeal to their interests. As one interviewee noted, making an application interesting and fun is an important motivation. Simplicity of interface and effortlessness are important characteristics which should be considered in motivating users.

“I would say that they must have interest in what they do, they must have fun while doing it. There are always different reasons at play here. For example my personal reason in participating in such contributions is that I love reading and books. I like building digital libraries. So adding metadata is just a task within the context of building a library. I don’t like to think that I am doing cataloguing or adding metadata” (Study2-Interviewee14).

As the interviewee noted, a social metadata application design should consider simple-to-use interfaces so as to make the task of the user effortlessly simple. The interviewee added that:

“[As a user] I don’t like metadata as a separate activity per se. In fact, I don’t love metadata but I understand that it is really important and is part of the big picture. It is

not like I do metadata for the love of it or for the standards such as MARC. I believe most people would add metadata if they understand the importance of a collection. For example, a lot of people in the Library of Congress Flickr Commons photo stream project, a lot of people consider it interesting to see the Library of Congress sharing their archival collections via Flickr and got attracted to it; they enjoy browsing through these pictures collection, and then they start adding tags. They are adding metadata without even realising that they are adding metadata. They had fun doing that” (Study2-Interviewee14).

The interviewees indicated that simplicity of interface and the fun of doing it at one’s leisure constitute key motivating factors for participation in socially-constructed metadata approaches. They suggested that metadata system interfaces be both interesting and appealing to users, such that users find it fun and exciting to spend time in creating metadata, whilst, at the same time adding value to their own search and browsing experience. Conversely, if a metadata system is considered as ‘work’ requiring concerted and immaculate effort, it is less likely to appeal to the greatest number of people as contributors. Thus, the design of social metadata systems should bear in mind what one would consider such elements of motivation.

8.5.3. Re-findability

As some interviewees identified, one of the motivations for users to add social metadata is the intention to re-find the information object at a later time. One interviewee observed that users add metadata primarily for themselves. The interviewee added, “The moment they add a tag, it implies that the next time they make a search they are more likely to use their own tags. So they are developing tools to help themselves. I think that would be enough motivation to make users describe library materials” (Study2-Interviewee17). Tagging helps users to re-find things for the taggers and also create a community around it hence increasing its utility. Another interviewee concurred, stating that the motivation for users is essentially personal, explaining “users want to organise and categorise their own preferences and collections in their own way. If you see LibraryThing, users simply tag it so that when they come back next time to the site, they can find materials. A simple tag helps to categorise” (Study2-Interviewee21). Users tag or bookmark a link for later retrieval. Even though each user does it for their own purpose, when such metadata is aggregated it plays a pivotal role in informing patterns of resource and metadata usage. Users also spend a great deal of time trying to find books and articles that they have borrowed or accessed before. Metadata, through tags, would thus greatly assist users in retrieving it again. Users are more likely to search using the tags that they have used to bookmark the resource.

8.5.4. Sense of Ownership

Users are considered to be better motivated in contributing metadata if they develop trust that they have control over the metadata they create. Ownership means that they have the ability to add, access, modify, and delete their own entries, based on their preferences. Socially-constructed metadata faces an open question regarding privacy and intellectual property rights of the metadata created. Who owns the metadata that comes through tagging, reviews,

ratings and recommendations? Can a university student, for example, have control of the metadata they created when leaving the institution? What about in cases where subscription to text-books and e-books expire?

One interviewee (Study2-Interviewee4) shared their concerns in relation to persistence of metadata over time in a socially-constructed metadata approach. In connection with this, one of the main concerns is ensuring the accessibility of socially-constructed metadata that is associated with external providers, where users annotate, tag, and write reviews of books, e.g. on Amazon's Kindle. As the interviewee (Study2-Interviewee4) further indicated "What happens to your annotations in the meantime? It would be quite interesting to see what will happen, when the first test case comes up". Similarly, the interviewee identified issues of ownership in cases where for instance what happens to metadata created by students after they graduated.

The sense of ownership and control of own entries by users would partly contribute to build confidence on the part of users on the usage of their contributions. As one interviewee (Study2-Interviewee14) noted: "I would say that if people understand that no one is stealing their contributions and if they think everyone benefits from their contribution that is already a key motivation". The interviewee added that users are less likely to contribute socially if they know their data will be used for commercial purposes. In their own words, the interviewee remarked thus: "If you are a for-profit company and ask people to freely contribute, people will not come because they know that such projects may end up benefiting companies, and not the community of people that they want to help. But, if the project is set up with the goal of benefiting the commons, people would value such projects. People want re-assurance that their contributions won't be exploited for purposes that they did not intend" (Study2-Interviewee14). The interviewee further suggested that "when people contribute, they expect something in return. It is collaboration 101. It seems there is a need for reciprocity, which can be direct or indirect. My understanding is that people contribute if they feel they are getting something back in return".

In relation to ownership, the issues of users' own terminologies in the creation and utilisation of metadata can motivate contribution, suggesting that librarians and metadata experts should not predetermine users' vocabularies, as users may use various terminologies. The interviewees thus indicated the importance of examining the various terms users use and consider the variations in terminology use.

8.5.5. Altruism and Reputation

The interviewees identified altruism (helping others) as a motivator for socially-constructed metadata contributions. The interviewees also believed that users who understand the importance of their contributions to the social good provide metadata in terms of reviews, recommendations and tags. One interviewee (Study2-Interviewee14) indicated, in order for users to be motivated by altruism, they may want to see commitment and clear purpose on the part of the information provider to collect, aggregate and utilise the metadata obtained from

users. The same interviewee also noted that some users are motivated to help others in order to build their own reputation as a social contributor. The interviewee remarked thus:

“In a university context, academics is mostly a reputation based system. Tenure and authority are important for academics. Another example is the open source software community; it is all built on reputation. If you build a good framework, people would come and generate every kind of metadata. You find such good reviews on GoodReads (social network based around books). I myself use such tools to compile a list of books I have read, and also attach metadata to it. In fact people can see my profile and look at what books I’ve read and the reviews I’ve written. In fact people then recommend me other new books to read and write reviews on. Some people commend me for my reviews” (Study2-Interviewee14).

Another interviewee added that prizes and competition can also be used to motivate users. Regular contributors are prized with a social status in the group or that their contributions are recognised. The interviewee added “right now social behaviour has changed. The number of friends you have on Facebook is one measurement of your popularity. So you can create such kind of incentives which motivates users to contribute” (Study2-Interviewee21).

8.5.6. Engaging with Users

Interviewees indicated that socially-constructed metadata approaches may create an opportunity for libraries to better engage with their users. Users, one interviewee (Study2-Interviewee9) noted, can play an important role in informing the library of their needs and preferences, thus involving them in the creation of metadata may constitute one of the mechanisms to solicit their requirements as well as collect relevant metadata about information objects through reviews, tagging and ratings.

User created metadata may provide an important clue to their information needs, use of terminologies and may help in creating an important indicator for libraries to areas where they should focus in organising their collections, including purchases and subscriptions to information objects. As one interviewee noted, socially-constructed metadata may help a library to better understand its user-base so that the library may cater to their needs (Study2-Interviewee12). One of the motivations to engage users using socially-constructed metadata approaches is to use users’ own social spaces to collect metadata. One interviewee (Study2-Interviewee1) believed their library has been incorporating users’ requirements from surveys, focus groups and feedback, others contended that libraries should be doing more than determining users’ requirements through suggestion boxes, thus suggesting the importance of directly engaging users (Study2-Interviewee2; Study2-Interviewee14; Study2-Interviewee21).

8.6. Quality Control

Quality, one interviewee (Study2-Interviewee21) emphasised, is an interesting but elusive concept. No one seems to object to the importance of ensuring quality; nonetheless, some interviewees were emphatic in asking ‘who determines the quality criteria?’ The interviewee

mentioned earlier stated that, “the main goal should be facilitating access. Libraries may have spent greatly on collections and thus may have acquired the [appropriate] resources, but if it is not accessed and used by users, there is no value to it”, thus pointing to the importance of measuring quality of metadata based on its effectiveness in making library information resources findable, accessible and usable. One of the major concerns associated with the implementation of socially-constructed metadata approaches is the issue of metadata control. Libraries are not yet sure how to manage socially-constructed metadata especially in view of its quality, reliability and persistence. One interviewee was particularly concerned about the challenges metadata quality control would pose in socially-constructed metadata approaches. The interviewee indicated the absence of time and resources for monitoring user metadata. The interviewee noted: “I would want that social tagging to be in addition to the high quality, on top of the high quality metadata record from the library. [But the library should not] devolve responsibility to the users” (Study2-Interviewee1).

Interviewees seem to agree that metadata control is important but their views on the degree of control vary widely. Some argued that quality should never be compromised at all, thus suggesting a strict librarian presence to vet and police the metadata entered by users, so that it is checked for accuracy and relevance before it is presented to other users. Others, however, contended that the moderation should be very minimal and that such moderation should come after the metadata has been made available to other users. The latter group of interviewees noted that the moderation should only affect metadata that may put the reputation of the institution into question and/or entries that are malicious and harmful. One interviewee (Study2-Interviewee12) stated that user-submitted metadata should be checked to protect the institution from legal and ethical ramifications that may be caused due to defamatory and derogatory content. The interviewee suggested that such content should be removed. The same interviewee suggested the importance of provenance metadata (7.4.3). Such provenance metadata would provide information with regard to the authority, context, history and ownership of metadata, thus informing users about the credibility and veracity of the information. In addition, the interviewee indicated that users themselves can help in providing provenance metadata. Some interviewees also suggested that the quality of metadata and assignment of weights (for example through rating the reviews) should be judged by the users themselves (Study2-Interviewee6; Study2-Interviewee8 and Study2-Interviewee14) (8.4.5.7).

Thus, metadata policing, most interviews believed, should follow the examples of Wikipedia. In Wikipedia, as one interviewee (Study2-Interviewee7) remarked, “People can put anything they want, but there is also some control over spam. I see no reason not to allow users putting in whatever information they want, but at the same time the library may want to prefer certain information over others”. Interviewees identified the use of user identification mechanisms as one mechanism to control malicious entries. One interviewee, who is also an active Wikipedia volunteer, indicated the importance of normalisation and laying a basic system structure to achieve a level of quality control. However, adding that such normalisation should be done both by experts and users. The interviewee noted, Wikipedia lays a basic structure, in which either the user ID or ISP number is used to identify the source of the

information. Dedicated volunteers and also any other user can then check the accuracy and veracity of the information entered in Wikipedia. With regard to this kind of monitoring, one interviewee said, “I think there is scope for policing to remove the unhelpful, the maliciously wrong, but when it comes to what a work is about, I would like the users to do it” (Study2-Interviewee12). Since socially-constructed metadata approaches are not yet fully implemented in libraries, the issues that surround them, including metadata quality control and mentoring have not yet been addressed properly. One interviewee mentioned that they want to see such strategies emerging from practise. The interviewee indicated the importance of striking a balance between “being open and allowing participation and at the same time defending the reputation [of the institution]”.

Whilst some librarians voiced their concerns on the likely negative effect of allowing users to tamper with library metadata (Study2-Interviewee1; Study2-Interviewee6; Study2-Interviewee18), others believe findability and accessibility of information resources should be placed higher in priority than mere quality of metadata records. One interviewee said that when “someone adds metadata, [it] doesn’t mean your quality has gone down” (Study2-Interviewee7). But another interviewee stated his concern:

“I have no objection to involving users in the cataloguing process, in terms of adding perhaps short reviews or key word tagging, provided that it is done in addition to the high quality metadata supplied by the library” (Study2-Interviewee1).

The interviewee did not completely reject socially constructed metadata approaches, instead emphasising that metadata quality is their utmost priority and concern, thus asserting that any socially-constructed metadata be vetted and quality assured by librarians. Further into the interview, the interviewee indicated that even though they saw some benefits to socially-constructed metadata, the cost of ensuring its quality would be prohibitively expensive.

On the other hand, other interviewees, especially metadata consultants and researchers, argued that the fear by some librarians about the lack of metadata quality in socially-constructed metadata is exaggerated. One interviewee contended that:

“It is certainly true that librarians have a better understanding of the rules, have a better understanding of why they are cataloguing but [user created metadata] is closer to the mind-set of a potential searcher than that of the librarian” (Study2-Interviewee12).

The notion of metadata quality, another interviewee argued is an elusive one, noting that what a librarian might consider low quality metadata might be relevant for a user who created it, adding that “people’s opinions about films is useful. Sometimes, they may not be right. They may be superfluous and silly, and so on, but we have to live with that, because that is us, as humans” (Study2-Interviewee8). In contrast, another interviewee strongly believes that librarians can accurately determine the quality of metadata and further asserted that quality should come first before quantity and that socially-constructed metadata should be done as additional data on top of a high quality, accurate and consistent metadata record set created by librarians. Some interviewees (Study2-Interviewee1; Study2-Interviewee15; Study2-

Interviewee18), believed that tidying-up socially-constructed metadata is crucial to ensure metadata quality, albeit their dire concerns from where such resources could be found.

However, other interviewees (Study2-Interviewee2; Study2-Interviewee9; & Study2-Interviewee14) indicated that user tags should be considered as it is, without being altered by librarians. One of the interviewees (Study2-Interviewee9) in particular noted that some user tags such as ‘nice’ should be understood in context and thus cannot be automatically dismissed by the librarian as irrelevant and low quality metadata, further asserting that “if 10,000 to 15,000 people socially tagged something as being ‘nice’, that might be the same as a four-star review in Amazon. And that information we know is very valuable. Not just to Amazon but it is valuable for social trending, indicating perhaps the book is worth reading” (Study2-Interviewee9). Another interviewee concurred, stating that metadata about metadata, as, for example, saying that a certain review or tag is ‘helpful’, is important for ascertaining its social value as more users get to use it. Similarly, ‘negative’ metadata which says that a certain tag or review is ‘unhelpful’ is also considered as equally important, indicating the possibility of social vetting/policing.

Some librarians, as one interviewee (Study2-Interviewee11) criticised, “tend to concentrate on perfection [...] and they spend their life trying to fix the edge cases, while for the rest of the world [the World Wide Web] ‘good enough’ is ‘good enough’”. The major limits of perfection are that librarians try to achieve it a priori. Without waiting for the perfect record, socially-constructed metadata approaches start from an incomplete record, and make it public so that the corrections, editions and quality control happens after-the-fact (post-hoc). As the interviewees indicated openness always brings issues of abuse of systems but as demonstrated in Wikipedia and similar crowd sourcing projects, such abuses are exceptions rather than norms. Thus, it is crucial to dwell on the norms rather than on the exceptions. Most importantly, the involvement of users in metadata creation does not preclude the active engagement of librarians and metadata experts. In connection to this, the interviewees cited the example of Wikipedia where the contributors are not always naïve and uninformed people but includes people who are deeply involved in their specific disciplines and are concerned about their subjects so that they not only add articles of good quality but also help to safeguard the veracity of entries by other contributors. Amidst all the pessimism on Wikipedia, it seems as if it has proved that socially-constructed encyclopaedias can be a possibility.

Whilst, as librarians variously indicated, metadata quality is an important concern and one that would potentially benefit users, it is crucial that the main goals of libraries should cater to address the information, research and education needs of their users by enhancing the findability and discoverability of information objects. Thus, libraries should perhaps forgo the notion of perfect and embrace the concept of usefulness, indicating thus the importance of involving users as metadata co-creators. However, this does not imply socially-constructed metadata approaches are without challenges and limitations.

8.7. Challenges to Implementing Socially-Constructed Metadata Approaches

The Focused Coding stage of the analysis resulted in the emergence of the following major challenges associated to socially-constructed metadata approaches (see Figure 8.13).

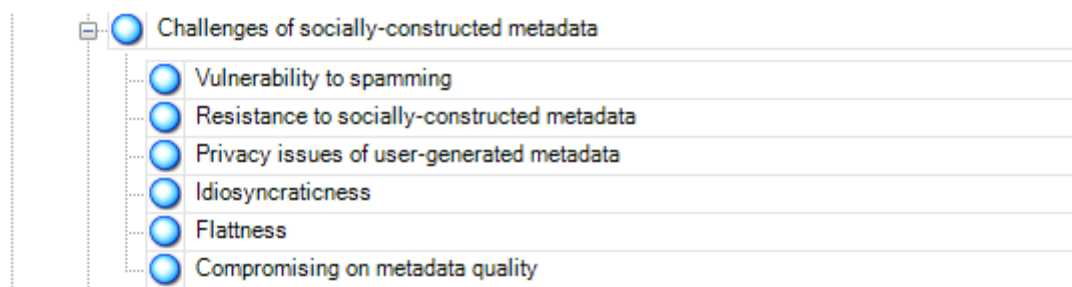


Figure 8.17: Challenges of Using Socially-Constructed Metadata Approaches

Whilst there is recognition of the importance of socially-constructed approaches for providing richer metadata descriptions, as the interviewees acknowledged, the practical use of such metadata features in libraries is neither well understood nor fully developed. There are a number of reasons for this. One of them is the relative novelty of socially-constructed approaches. Hence, as one interviewee remarked “There is a lot of experimentation in this sphere of social media areas” (Study2-Interviewee13). Another reason is the fact that the notable early attempts at social tagging and recommendation services are owned by commercial entities, making integration with libraries difficult. A third major reason is the proprietary and rigid nature of Library Management Systems (LMS) where adding new metadata features into the cataloguing module becomes rather expensive, partly because such LMS operate on low profit margins, thus do not encourage developers to add new features.

There are now new attempts to add social tagging into some library systems. However, as one interviewee noted “Primo based new OPAC system allows tagging, but the take up of tagging has not been tremendous. We have not had a huge amount of people rushing into that kind of thing - some have begun to add tags, more in the last few months, but it still hasn’t really taken off. What, therefore, we have is a quite patchy coverage, in terms of use of tagging. But, pulling the tags, from a variety of different institutions, might immediately give people a certain level of critical mass” (Study2-Interviewee4). The attempts in libraries thus far are as add-on and very peripheral and there is no serious consideration of socially-constructed metadata approaches in the main functions of library metadata.

Even when it is implemented, socially-constructed metadata approaches, partly attributed to it being done post-hoc, is vulnerable to spamming by malicious users. Since anyone can potentially add metadata, the enormity of the metadata would make it cumbersome to manage for libraries. In addition, this type of metadata has no hierarchical (synonym/homonym) control, thus its flatness may rob its contextual significance (9.3 for some of the solutions to this problem).

8.8. Conclusions

In contrast to a priori metadata (7.3.3) which characterises the metadata created using standards-based (expert-created) metadata approaches, post-hoc metadata emerged as a category (8.2) which depicts the co-creation of metadata by library users. In addition, the Focused Coding of interview data resulted in the emergence of categories which are pertinent in the consideration of socially-constructed metadata approaches. These categories include: platform for metadata creation (8.4.1), users as proactive metadata co-creators (8.4.2), metadata diversity (8.4.3), metadata scalability (8.4.4) and collective metadata intelligence (8.4.5). Figure (8.18) shows the relationship between the various Categories of socially-constructed metadata approaches.

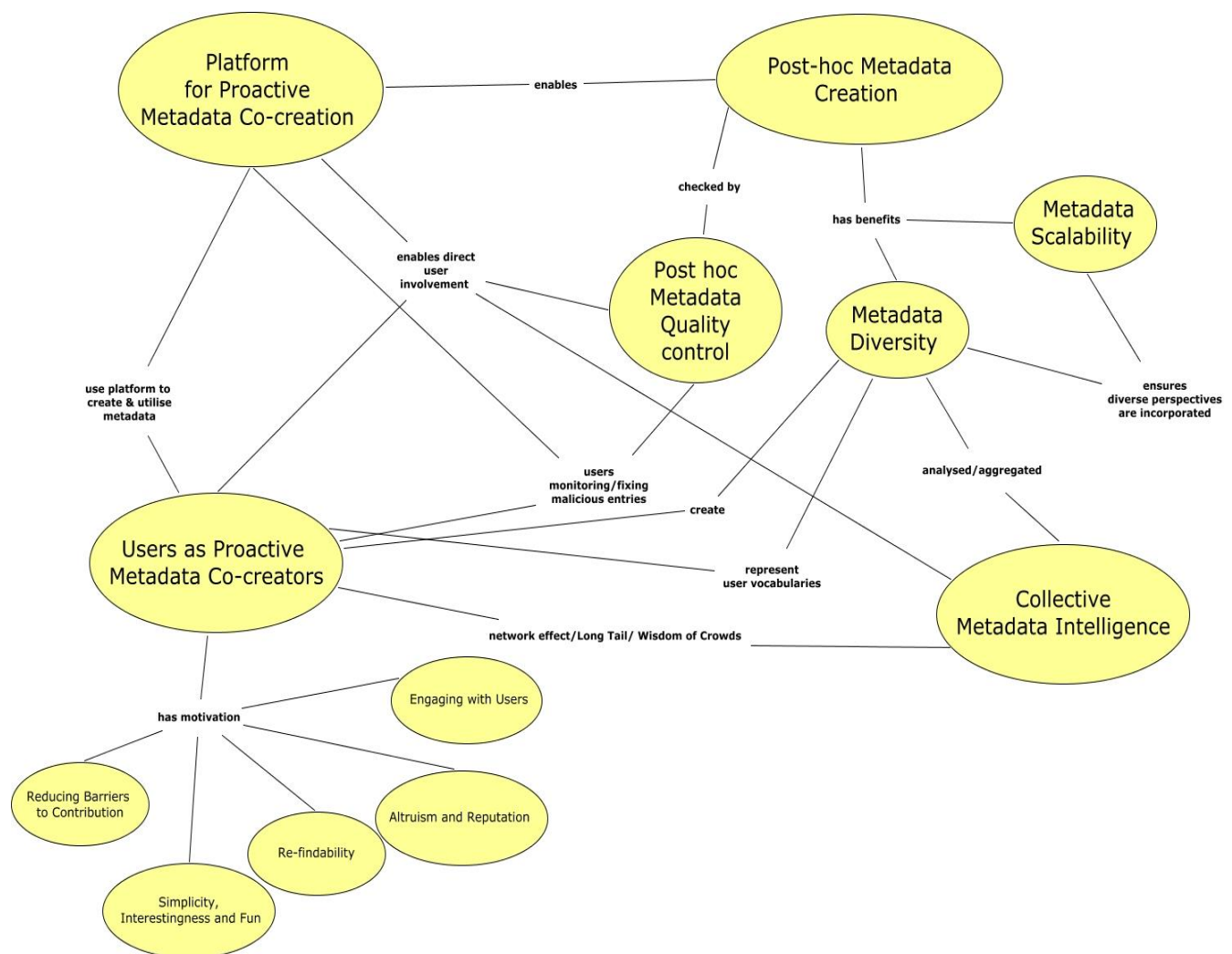


Figure 8.18: Relationship between Categories of Socially-Constructed Metadata Approaches

The analysis also identified the challenges of socially-constructed metadata approaches including its flat structure, vulnerability to spamming, and quality control issues. Thus, it is indicated, in order to optimally benefit from socially-constructed metadata approaches, these challenges need to be addressed (8.7). Whilst some of the challenges can be addressed by involving users themselves in metadata management and quality control functions, the importance of metadata structure that is present in standards-based metadata approaches

(7.4.1) is suggested. Such combined use, it is indicated, would help provide rich metadata descriptions.

The findings reveal that contemporary standards-based metadata approaches, conforming to read-only (Web 1.0) paradigm, exhibit serious limitations in adequately satisfying users' needs. Furthermore, as the size of collections in libraries grows, these principles and standards fail to scale when describing and/or enriching descriptions of information objects. To alleviate both challenges, the analyses demonstrated the need for harnessing the Web 2.0 paradigm that provides a technological and social platform, wherein users can be accorded the role of proactive co-creators, rather than being considered passive consumers. It is also recommended that libraries move from what is often-referred to as a user-centred mentality to one that is user-driven, as the latter would better enhance the realisation of the principle of user convenience (2.7).

Thus the creation and management of metadata should no longer be the sole prerogative of an expert librarian; instead, the librarian (expert) would provide a set of metadata entries about an information object, which can further be enriched (Chapter Nine) by contributing users through conversations (tags, ratings, reviews and recommendations). Whilst the findings revealed that contemporary metadata is still firmly in the grasp of librarians; the possibility of augmenting standards-based metadata with the social-space of metadata has also been raised. The identification of the Categories in both standards-based (Chapter Seven) and socially-constructed metadata approaches (Chapter Eight) led to the emergence of the Core Category of metadata enriching, which is discussed in the following chapter (Chapter Nine).

Chapter Nine: The Core Category of Metadata Enriching

9.1. Towards a Mixed Metadata Approach

The data analysis in Chapters Seven and Eight revealed the strengths and limitations of both standards-based and socially-constructed metadata approaches. In Chapter Seven, the important Categories identified and developed included metadata structure, granularity, provenance, quality, simplicity and interoperability. Similarly, the data analysis in Chapter Eight also indicated the potential role of socially-constructed metadata approaches, with their Categories being identified as an enabling platform for proactive metadata co-creation, role of users as proactive metadata co-creators, metadata diversity that resulted from such co-creation and collective metadata intelligence that can be garnered from it. The data analysis showed that the two metadata approaches can better be utilised complementarily (see Memo in Figure 9.1).

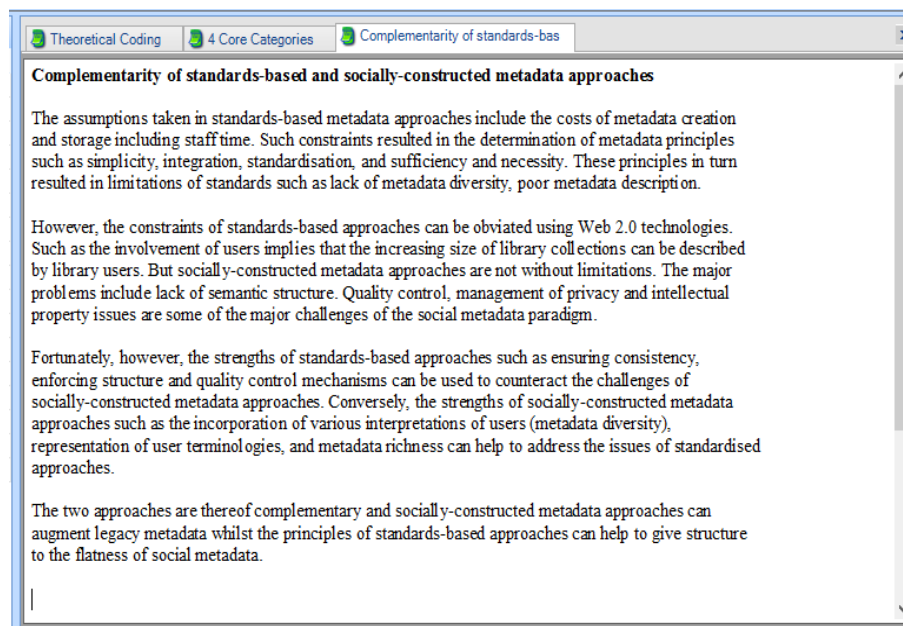


Figure 9.1: Memo on the Complementarity of Standards-based versus Socially-Constructed Metadata Approaches

The interviewees noted that socially-constructed metadata approaches are not mutually exclusive with standards-based metadata ones. One interviewee (Study2-Interviewee14) considered standards-based and Web 2.0 approaches as complementary, noting that, such mixed use should take into account the nature and purpose of projects. Most interviewees agreed that socially-constructed metadata approaches result in richer metadata but indicated that such metadata is predominately flat, and lacks semantic structure. It is at this stage, as interviewees indicated, that standards-based metadata approaches can provide the requisite hierarchical and semantic structure. One interviewee believed that mechanisms should be

sought to implement the two approaches in concert. The interviewee noted that one does not necessarily overtake the other, but solutions should cater at the level of data integration (Study2-Interviewee8). The diversity of interpretations and subjective descriptions of information objects that users may provide, the interviewees acknowledged, can prove to be a valuable set of metadata. One interviewee mentioned their wish to see the two approaches implemented in their library, so that the present state of the metadata and its presentation to users is improved (Study2-Interviewee2). Another interviewee (Study2-Interviewee12) highlighted the importance of embracing the two approaches. Nonetheless, the interviewee admitted that there exists at present a tension between these approaches: “I think it will take time for everything to be switched over and changed. But I don’t see everyone dropping MARC or dropping traditional systems overnight, nor should they. They don’t have to. So they can co-exist” (Study2-Interviewee12).

Supporting a mixed metadata approach, one interviewee identified two reasons for using socially-constructed metadata approaches. These are, firstly, users are capable of producing good metadata. To support this, the interviewee, a former cataloguer, saw a common principle in producing metadata for libraries, without getting paid for their effort and believed that there are like-minded people who would similarly volunteer creating and maintaining metadata. The second reason, the interviewee indicated arises from the use of the “Wisdom of Crowds” (8.4.5.6) and crowd sourcing approaches for metadata creation and maintenance (8.4.1.5). One interviewee believed that the prerogative should be left to the library as to whether certain metadata is preferred over another, so that the library makes the final decision as to the presentation of the metadata to users (Study2-Interviewee7). Another interviewee emphasised that both approaches can be used together, as long as it helps the information user to find and determine the relevancy of an information object (Study2-Interviewee11).

The limitations and challenges of socially-constructed metadata that were described by interviewees are mostly structural (8.7), which can partly be addressed through standards, thus, the Category of Structure (7.4.1) of standards can be harnessed. Structuring socially-constructed metadata is one of the ways for optimising the aggregated potential of socially-constructed metadata. Of course, the standards need to retrofit the nature of social metadata. It, for example, should be less stringent and constraining for contributors. Otherwise, users would be discouraged. From a usability point of view, the structures put in place should be made user-friendly, memorable and enjoyable. Even more crucial is the ultimate goal of the metadata, which can substantially affect the motivation of contributors.

“Flickr is quite an example of folksonomy where people puts tags to their pictures, so it is important to give users some sort of system, such as close matches of new tags with existing tags, a kind of ‘did you mean this tag’, if not, it could be added. You also need some sort of mechanism if you need a browsable set of subject terms, find a way where a new term can slot in an existing hierarchy of tags/terms [the main challenge is] to design a system which can cope with these metadata and rationalise it” (Study2-Interviewee10).

As interviewees suggested, standards-based metadata can serve as a stable structure that socially-constructed metadata can be mapped to. Based on its context, related tags can be collocated and mapped into a thesauri-like ontology where, as some users suggested, Linked Data technologies could help. One interviewee (Study2-Interviewee4) for instance suggested: “A good hybrid of all these might be, when people are given the opportunity to tag items [using existing list of terms, whilst still allowing] people to put their words” (Study2-Interviewee4).

In connection with this, interviewees suggested the importance of subject headings and controlled vocabularies (see 2.10.4; 7.4.1). It is indicated that these controlled vocabularies can be improved by users as they are allowed to contribute in a mixed metadata approach, where users enhance and enrich metadata created by experts.

It is indicated that once tags are aggregated, patterns of similarity and semantic relatedness are established using algorithmic means. Socially-constructed terminologies can, for example, if an institution decides, serve as suggestions of terms for user tags. Such suggestions should be carefully implemented as users should still be free to add their own tags. Tag suggestions, if implemented, should not limit and constrain users from adding new tags. Interviewees also indicated that, as the size and diversity of social metadata is expected to be large, it is important that automatic and algorithmic approaches are used rather than manual labour.

It is a matter of institutional policy whether to tidy up metadata such as misspellings and obvious errors. This is a fluid area since in some circumstances; libraries may simply want to leave the errors as they are since users are also more likely to search using the same keywords. Similarly, the issues of idiosyncratic tags such as ‘to be read’, ‘mum’s gift’, and many other personal tags could be hidden from public view (12.2).

Some interviewees suggested the use of Linked Data to structure socially-constructed metadata. But they highlighted the importance of mixing and merging various metadata from disparate sources including controlled vocabularies, authority lists and multi-lingual translations. Some of the light weight ontologies suggested included that of schema.org (Study2-Interviewee19). It is indicated that, unlike traditional controlled vocabularies, these ontologies should be populated both by experts and users. One interviewee (Study2-Interviewee20) noted that “librarians can have the role of editing and supporting ontology building but I really think this can be done in partnership with the user. I don’t even see an alternative to this. I see convergence”. Another interviewee, however, saw limitation in the use of ontologies:

“Yes, you could build ontology but ontologies are difficult to develop and implement. Ontologies are inflexible and one error causes the whole system to disrupt. But if you think Wikipedia, it is not a thesaurus of course, but has a few simple systems for normalising words. There is a minimal level of synonym/homonym control implemented in Wikipedia [so as to] disambiguate Wikipedia entries. If you search

for the term ‘law’ in Wikipedia, it is disambiguated to the various categories such as law in the legal system context or a name of a person” (Study2-Interviewee14).

It is indicated that users should be given control over their own metadata. Thus, they can add, edit, delete and manage their own metadata, whilst the metadata system can aggregate and harness individually created metadata. So the system should be able to gather and aggregate individual metadata so as to optimise individual and collective user experiences. Noting the possibility of integrating the two metadata approaches, one interviewee (Study2-Interviewee11) indicated that mappings between vocabularies of the two can be made behind the scenes.

Another interviewee noted that

“I think they can co-exist. But whether you combine them into one, which is quite a tricky question. At the moment, I mean I suppose things like Google work is that they utilise both but they give higher rankings to certain things. The search tool can do that. You can boost certain metadata. Making it appear first in the result. So you could have standardised metadata appear first” (Study2-Interviewee6).

The assumptions taken in standards-based metadata approaches include the costs of metadata creation and storage including staff time. Such constraints resulted in the determination of metadata principles, such as the principle of representation (2.8), the principle of sufficiency and necessity (metadata simplicity) (2.6) and the principle of standardisation (2.9). Unfortunately however, these categories of standards-based metadata approaches have serious limitations when it comes to enriching metadata. The more standards place stringent constraints to fit a predetermined schema, the more these strictures limit the size of metadata entries as such rules only allow very restricted number and type of values, not the least the values should be created by metadata experts who assume some level of authority. Thus these requirements could become constraints for scalability. In addition, since for quality and interoperability concerns, only authoritative sources of controlled vocabularies, which come from trusted sources, are allowed, the vocabulary of users is missing, thus also affecting metadata diversity. Mixed metadata approaches that cater to enrich metadata better matches interviewees’ concerns and requirements to see a metadata system that places users’ as proactive co-creators rather than passive consumers of metadata.

From the data analysis, the case for the interdependence between the two metadata approaches is made in consideration of the importance of a priori metadata so users can subsequently be able to create post-hoc metadata entries. Conversely, post-hoc metadata is essential to augment a priori metadata. Hence, such interdependence and combined use of metadata approaches give rise to an emerging metadata approach called a **Mixed Metadata Approach**.

During Theoretical Coding, the focus was in identifying an overarching Core Category that better represents the underlying purpose of combining the two metadata approaches. For this purpose, the researcher used memo writing and reflexive conceptualisation. The Core Category does not necessarily subsume the Categories in a hierarchical (top-down)

relationship. But it is important that the Core Category is grounded in the data that helped to identify the Categories in the first place, thus the Theoretical Coding results in Core Categories through an iterative process. It is important to note that as the data analysis progresses from Open Coding (Open Codes) to Focused Coding (Categories) and subsequently to Theoretical Coding (Core Categories), the level of abstraction and conceptualisation of the research findings also increases.

As a result of complementarities between the two approaches, metadata can easily be aggregated and analysed, enabling one to harness the power of collective metadata intelligence. Collective metadata intelligence can be realised through the provision of a platform for metadata co-creation, exploiting the Long Tail of metadata phenomenon, leveraging network effects, and mobilising the “Wisdom of Crowds”. Furthermore, metadata diversity can be quality-controlled, not only by librarians, but also through the involvement of users, which would result in a self-healing metadata system where malicious entries can be reported, edited and regulated by the users themselves. In addition, provenance metadata can be appended to socially-constructed metadata so that its source, authority, context and history can be determined, thus informing users about the authenticity and integrity of the whole.

Whilst each metadata approach has its own strengths, it also exhibits inherent limitations:

- Standards-based metadata approaches are susceptible to difficulties associated with metadata scalability, duplication of effort, metadata silos, slowness in accommodating changes, and failure to adequately engage with users;
- On the other hand, socially-constructed metadata approaches are limited by issues emanating from lack of structure (flatness), idiosyncratic entries, as well as the absence of quality control mechanisms for excluding undesirable activities, such as spamming and malicious metadata content.

The main strength of socially-constructed metadata approaches lies in their potential for reducing the marginal cost of metadata generation, by having users participate in metadata co-creation. Whereas, in standards-based metadata approaches, cost minimisation was rightfully considered a major justification for establishing the Principle of Sufficiency and Necessity, whereby metadata experts justified limiting the number of metadata elements to a selected few, in conformance with Occam’s razor principle (see 2.6). However, this assumption has now become relatively irrelevant, as the incremental cost of metadata creation is low in socially-constructed metadata approaches. Hence, some current metadata principles underpinning standards-based practices should be re-conceptualised. However, such re-conceptualisation should be based on the assumption that standards-based and socially-constructed metadata approaches are complementary to each other, rather than presuming that the latter is a replacement for (or contender of) the former.

The strengths of standards-based metadata approaches would be able to compensate for the limitations of socially-constructed metadata and vice versa, whereby each complements the other. Thus, the absence of structure in socially-constructed metadata can be obviated with the structure inherent in standards-based ones. On the other hand, strengths of socially-

constructed metadata, such as the ability to involve users as metadata co-creators, and the diversity of the resultant metadata, obviate standards-based metadata approaches' shortcomings such as scalability and lack of user engagement. Hence, a mixed metadata approach that accepts the complementarities of each to the other is strongly suggested.

Using the emergence of a mixed metadata approach as a spring board, the subsequent stages of constructivist grounded theory analysis Theoretical Coding resulted in the emergence of the first Core Category, namely Metadata Enriching. The Core Category of metadata enriching conceptually subsumed the Categories of the Focused Coding stage of data analysis in relation to the perspectives of LIS professionals on the role of socially-constructed approaches in current standards-based library metadata.

The overarching aim of the Categories of both metadata approaches is that of Metadata Enriching. It refers to the process of enriching existing a priori metadata with post-hoc (socially-constructed) metadata. Memo writing also helped to develop the Core Category. In accordance with the Grounded Theory Method, the gerund form of the phrase "metadata enriching" is preferred over "metadata richness", since the former indicates a continuous process of user metadata co-creation.

9.2. The Core Category of Metadata Enriching

Metadata enriching indicates the diversity (8.4.3), granularity (7.4.2) and breadth of metadata in its description of information objects (see Figure 9.2)

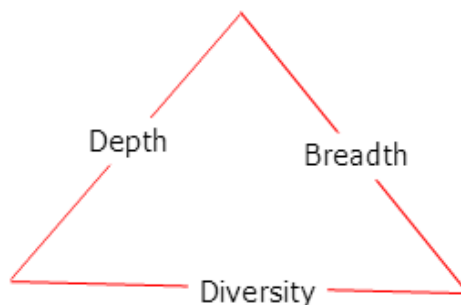


Figure 9.2: Dimensions of Metadata Enriching

9.2.1. Metadata Diversity

As discussed in Chapter Eight (8.4.3), metadata diversity deals with the representation of the various perspectives/viewpoints of different groups of users, whether this pertains to age, educational level, and socio-cultural, economic, or other context. The use of socially-constructed metadata approaches supports the incorporation of various interpretations of information objects. Existing standards-based metadata can be enriched through the incorporation of multiple viewpoints (8.4.3). Such metadata diversity should be structured (7.4.1; 8.4.5.4), affixed with provenance (7.4.3; 8.4.5.8), ensured of an acceptable level of quality (7.4.4; 8.6) and made interoperable (7.4.6). In other words, in order to optimally benefit from the metadata diversity of socially-constructed metadata approaches, the strengths of standards-based metadata approaches are indispensable.

Metadata diversity is possible only through the incorporation of diverse interpretations. It involves the Long Tail (8.4.5.2), the network effect (8.4.5.5) of metadata and through the Wisdom of the Crowds (8.4.5.6), thus indicating that metadata diversity is not merely the collection of metadata but the aggregated and analysis of metadata. In other words, metadata diversity (8.4.3) and collective metadata intelligence (8.4.5) are highly interdependent.

In addition, a mixed metadata approach generally presumes metadata diversity rather than metadata simplicity. However, simplicity is important in presenting it to users (see 8.5.2). This is a major paradigm shift from the principle of sufficiency and necessity (discussed in 2.6). A metadata description enriched with a diversity of interpretations indicates that for each metadata field, there are multiple values showing the diversity inherent in the interpretation of information objects by users. It is important to note that the value of metadata enriching does not imply a mere abundance of redundant and irrelevant metadata, but the prevalence of a diversity of relevant interpretations.

9.2.2. Metadata Granularity

As discussed in Chapter Seven (7.4.2), metadata granularity is also suggested as one of the dimensions of metadata enriching as it refers to the level of detail in the description of information objects. The granularity (depth) of metadata refers to the level of specificity and detail in description. Granularity can be looked at from the point of view of the structural components (e.g. chapters and sub-categories) of an information object.

It indicates whether an information object is described with metadata at a book, chapter, or section level. Granularity is discussed in the metadata literature under the typology of structural metadata. Such metadata helps to bring together various component parts of an object through the description and linking of the component parts (Zeng & Qin, 2008). The interviewees indicated that METS is the dominant standard in this regard. The focus here, however, is not as such, on METS but the implications of granularity for metadata enriching. In relation to the importance of granularity in metadata, one interviewee (Study2-Interviewee4) stated that “when a book is digitised, it could actually consist of hundreds of JPEGs, each printed page with one image”. The interviewee identified the METS standard as a useful format for encoding relationships between component parts of information objects. The interviewee added that greater degree of granularity and detail requires greater degree of effort in both metadata design and implementation. Granularity is therefore an important criterion to evaluate the depth of coverage of metadata, which in turn contributes towards metadata enriching. In the context of a mixed metadata approach, it is mainly the standards-based metadata that would help to provide the required granular structure, whilst the actual metadata values could come both from a priori (expert-created) and post-hoc (socially-constructed) metadata. A metadata system with good granularity should include componentised metadata fields, which would enable the description of an information object at the required level of depth or detail.

9.2.3. Metadata Breadth

In addition to metadata diversity and metadata granularity, the Core Category metadata enriching has a third dimension, i.e. breadth. Breadth is related to the extent and amount of metadata associated with an information object. In connection with metadata breadth, in the preliminary study (5.5), interviewees stated that the MARC standard provides more metadata breadth. Their assertion is based on the number of fields available in MARC (it has about 100 metadata fields (tags)), in stark contrast to Dublin Core (which has fifteen fields only). On the other hand, the participants welcomed the simplicity of DC.

9.3. Separation of Metadata Creation and Presentation Processes

The separation of metadata creation and presentation emerged as an important consideration. For instance, one interviewee noted that:

“I think there is a need for richness behind the scenes, because that richness can then be used to power quite powerful applications. It could be used to derive personalisation; it could also be used to derive flexibility and choice. The problem comes when you try to display that richness to the user. A number of OPACs, for quite a number of years, will either show you very basic information such as title, author, date of publication and that was about it. And if you asked for more than that you, as a member of the public, would be shown a MARC record” (Study2-Interviewee12).

In regard to this idea of simplicity, a distinction should be made between what is presented to the user as metadata (Chapter Twelve) and what goes on behind the scenes to describe information objects. Sadly, at present, the notion of metadata simplicity is misapplied. Reducing the number of metadata elements robs information objects of the descriptive, contextual, provenance, technical and administrative information associated with them. To put it differently, incorporating diversity of perspectives, and employing a richer set of metadata elements (fields) may maximise metadata enriching. It is therefore important that a distinction be made between interface simplicity and metadata enriching, as the first implies a technical solution whilst the latter denotes a semantic concern.

9.4. Decreasing Marginal Cost of Metadata Creation and Management

The principle of sufficiency and necessity (IFLA, 2009; Svenonius, 2000) rightly takes into consideration the cost of creating, storing and managing metadata. This is as it should be, as metadata creation, storage and maintenance have significant cost implications. However, approaches such as metadata crowd sourcing are indicative of a reduction in the marginal cost of metadata as the cost, in terms of time, to create records is distributed among users. Storage costs of libraries are also being reduced, as new approaches such as cloud storage are coming onto the scene. With a mixed metadata approach, the categories of socially-constructed metadata approaches, such as platforms for proactive metadata co-creation (8.4.1), proactive involvement of users (8.4.2), network effect of metadata (8.4.5.5.) and user motivations to participate in metadata creation (8.5), all indicate a reduction in the marginal cost of metadata creation.

Contemporary metadata standards are underpinned by the principle of sufficiency and necessity. The resultant effect of metadata simplicity is that the description becomes superficial, lacking breadth, depth and diversity. In other words, some current standards such as Dublin Core lack granular structure and detail not necessarily due to technical limitations but as a result of the principles that underpin the metadata standards and their implementation.

9.5. Conclusions

The Constructivist Grounded Theory Coding resulted in the emergence of a mixed metadata approach, which implies the inclusion of the multiple interpretations by users (post-hoc metadata) of information objects so as to augment or improve metadata created by library experts (a priori metadata). It is indicated that the strengths of standards-based practices such as granularity, provenance, simplicity, quality and interoperability can be harnessed to make sense of the resultant multi-faceted metadata of socially-constructed metadata.

In the light of mixed metadata approaches where socially-constructed metadata approaches can be implemented by involving users in the metadata creation and management process, the principle of sufficiency and necessity (2.6) can be re-considered. It can be argued that the concern of metadata cost minimisation, which underpins the principle of sufficiency and necessity, is partly obviated through the use of Web 2.0 paradigm and its emerging technological trends.

The results of the Focused Coding data analysis (see Chapter Seven and Eight) indicated the importance of adopting a mixed approach where metadata created by experts (standards-based metadata) can be continually enriched by users (see Figure 9.3).

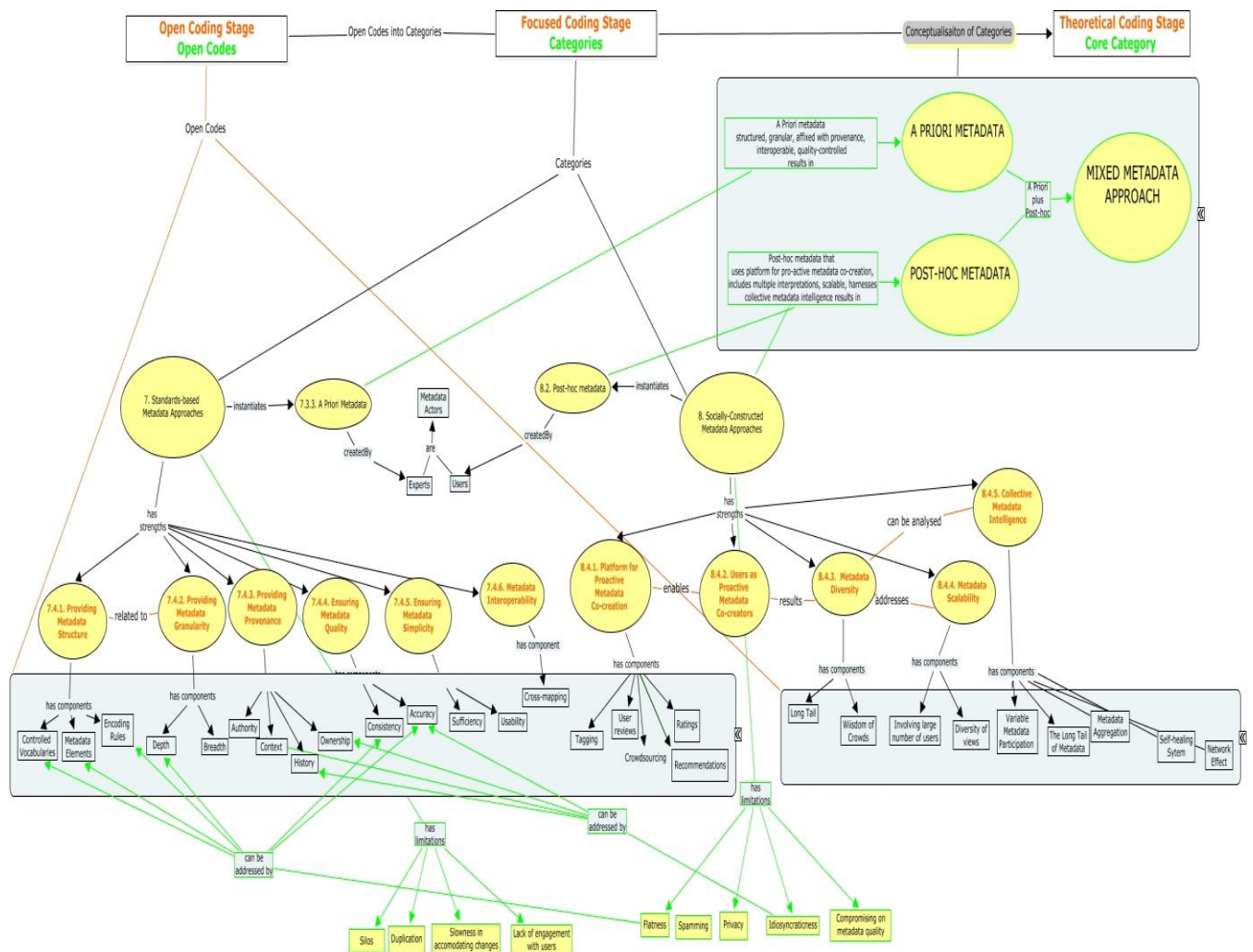


Figure 9.3: Towards a Mixed Metadata Approach

After conducting the main study (Study Two), in this thesis, a closer examination of both standards-based and socially-constructed metadata approaches revealed that a priori (expert created) (7.3.3) and post-hoc (socially-constructed) (8.2) metadata should be considered as complementary and interdependent rather than in isolation. The impetus for socially-constructed metadata approaches is the underlying conceptual and technical infrastructure of the Web 2.0 paradigm. The paradigm presents a new opportunity for libraries to implement a viable platform for metadata co-creation (8.4.1). Decreasing marginal costs of metadata generation, through distributed inputs, as well as the progressive reduction in metadata storage costs provide incentives for libraries to embrace metadata co-creation and metadata diversity (8.4.3). Similarly, the ability to aggregate and analyse socially-constructed metadata results in collective metadata intelligence (8.4.5). Important theoretical concepts for consideration under collective metadata intelligence included variable level of participation (8.4.5.1), the Long Tail of metadata (8.4.5.2), metadata aggregation (8.4.5.3), structuring social metadata (8.4.5.4), the network effect (8.4.5.5), the Wisdom of Crowds (8.4.5.6), self-healing systems (8.4.5.7), and affixing provenance metadata (8.4.5.8). In addition, in order to encourage metadata co-creation, incentives are crucial (8.5). Some of the motivating factors

identified include reducing barriers to contribution (8.5.1), simplicity, interestingness and fun (8.5.2), re-findability (8.5.3), sense of ownership (8.5.4), and altruism and reputation (8.5.5).

Using the emergence of a mixed metadata approach as a spring board, further data collection and analysis resulted in the emergence of the first Core Category, called metadata enriching. Metadata enriching is thus an overarching Core Category, and can be contrasted with the lack of adequate metadata description of an information object, something interviewees ascribed to the limitations inherent in standards-based metadata approaches. The Category of metadata enriching is chiefly a qualitative concern, focusing on the depth, breadth and diversity of metadata that describes the essence of an information object. The use of gerund in ‘metadata enriching’ indicates that it is an on-going process of enhancing the description of an information object. This is in contrast to current metadata principles which tended to result in metadata created once by a metadata expert.

The principle of metadata enriching suggests that the Web 2.0 paradigm offers a platform for users to enhance a priori metadata (which is created by metadata experts) with post-hoc metadata through tagging, reviews, ratings and recommendations. Both standards-based and socially-constructed metadata approaches play in concert, each offering what it does best. Standards give structure to the socially-constructed metadata, whereas the latter in return offers the diversity of descriptions that is obtained from users. Finally, it is important to point out that, consideration of metadata enriching has other important implications, one of which is that metadata that is enriched adds value to both institutions and users, if it is effectively linked, which is the topic of the next chapter.

Chapter Ten: The Core Category of Metadata Linking

10.1. Chapter Overview

In Chapter Nine, the Theoretical Coding stage of data analysis revealed the Core Category of metadata enriching (9.2) as an underlying overarching principle that lies beneath the overall aim of adopting a mixed metadata approach (9.1), which combines the categories of both standards-based and socially-constructed metadata approaches. Similarly, subsequent Theoretical Coding of data analysis revealed the importance of linking metadata that is enriched through the combined use of standards-based and socially-constructed metadata approaches. Thus, through closer iterative examination of the interviewee data, memo writing and by conducting additional interviewee data interviews the Core Category of metadata linking emerged. Further analysis of the Core Category of metadata linking also resulted in the identification of its Categories. These Categories are used to structure the discussion. In addition, in accordance with the Constructivist Grounded Theory Method, memos were also written to support the analysis with self-reflexive discussions.

As this chapter discusses, the Core Category of metadata enriching (see Chapter Nine) can only be optimally utilised if it is linked in-house (internal linking - within a particular library metadata set) as well as to external metadata sources (external linking). Linking metadata within and outside the library domain is deemed crucial for metadata enriching. A single institution, of whatever size and resources, cannot continually enrich metadata on its own. Linking can thus be used for further enriching, preferably in an open environment rather than in a closed one (silos). Linking is a powerful means for creating seamless connections between disparate sources of data. Links also make library metadata visible. However, the interviewees noted that current library metadata linking is not well developed. The interviewees attributed the challenges to both institutional (intellectual property rights and metadata quality concerns) and technological (metadata granularity, structure and formats) factors.

Links are considered the most important enabling components of metadata. The analysis of interview data discussed hereafter highlights the importance of linking technologies. Interviewees were in particular keen on the use of Linked Data (RDF) for representing and linking library metadata. It is important to note here that various interviewees identified linking as an essential part of metadata (see 6.7; 7.3.1 & 7.4.2).

10.2. Contemporary Library Standards and Metadata Linking

The incompatibility of current library formats to easily cross-link, among metadata values within and to outside libraries is considered a major shortcoming. As the interviewees indicated existing metadata approaches did not foresee the need for a linking structure, one that could be easily harnessed by simply pointing to other metadata values. One of the notable metadata standards interviewees cited for failing to adequately enabling a metadata link structure is MARC. As discussed in Chapter Seven (7.6), current metadata standards,

such as MARC, consist of structures that do not favour linking among metadata values or making connections to external sources. This was partly due to the fact that MARC was developed in the 1970s, before the invention of the World Wide Web. MARC, one interviewee noted, was an attempt to computerise the contents of card catalogues, hampering thereafter the implementation or introduction of a link structure (Study2-Interviewee11). The interviewee recalled that, in its early days of the 1970s, libraries used to display metadata embedded in MARC on character-based dumb-terminals. Many interviewees thought that libraries are to be praised for their early recognition of the value of computers and adopting them.

However, as one interviewee (Study2-Interviewee7) noted, “libraries started creating metadata long before other people did. They already had their concepts developed. Libraries have been creating metadata for 150 years, so libraries tend to have concepts that predate data processing. So the standards were developed to make use of those concepts” (Study2-Interviewee7). Figure 10.1 is a simple attempt illustrating the current state of MARC-formatted records, where it is evident that there is poor structural support for metadata linking between databases, among records or between metadata elements.

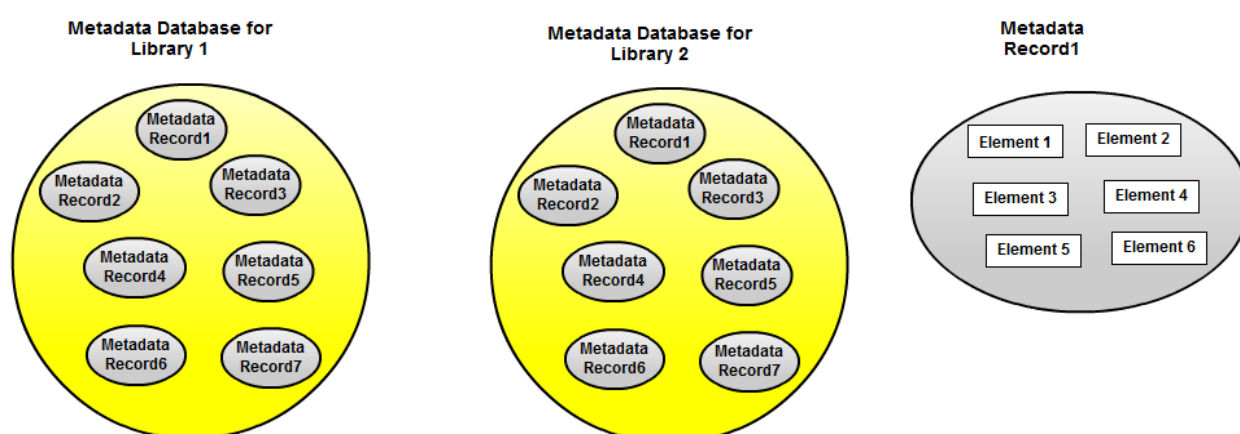


Figure 10.1: Metadata that is not Linked

The figure illustrates the absence of linking within the metadata system (compare to Figure 10.5). One interviewee pointed out the inherent structural challenges as well as the obsolescence of the formats found in contemporary library metadata standards (Study2-Interviewee11). The interviewee noted that “the library management community is highly concerned about the way they store and disseminate their data and believe that Linked Data is probably the way forward. The rest of the library community doesn’t know, doesn’t understand, quite frankly, they wish things didn’t change” (Study2-Interviewee11). Linking has powerful implications. David Weinberger (Weinberger, personal communication, December 12, 2012) had the following to say:

“I can imagine a circumstance where a library just doesn’t want to provide linking to external sources. But I think that is an unusual circumstance. The content of the

library has more value generally when it is linked to other ideas and information. This is not true only for libraries; it is true for all other information providers. It is not so much that information wants to be free; it is that information wants to be linked. It wants to find other information. To cut that off seems to me to want to value the authority of the information over its utility. When you want to know something, almost in every case, you don't want to be finished with it. Sometimes you do but in many cases libraries should and need to stimulate our curiosities to the next piece and hope that next piece also leads us to more curiosity, to guide our pursuit of curiosity. There is no reason why that should occur within the library collection or the library walls and so there seems to me a natural and almost an inevitable desire to pass that boundary. That boundary is artificial. So either you encourage that and make the linking better or pretend that it is not going to happen”.

Interviewees indicated, there is recognition now that, as requirements for web-based scalable metadata formats grow, such library-centric formats are becoming obsolete when it comes to supporting scalable linking. However, there is a need for data convergence, not only among libraries but also across other information sources and data providers, thus the need for adopting new formats becomes all the more relevant. Most interviewees discussed the current status of library metadata formats in comparison to XML and RDF. These new formats were considered more appropriate for representing library metadata. Interviewees acknowledged the importance of re-designing library metadata, in view of the need for linking in-house and with other data sources. The interviewees emphasised the importance of replacing these formats with flexible and scalable ones. However, it was evident that the data represented using these old formats needed to be compatible with the new formats.

One interviewee (Study2-Interviewee9) pointed out that library standards have been characterised by linear progression rather than disruptive technologies. The interviewee stated that most of the developments in library catalogues have been rather straight forward. The linear progression, the interviewee believed, has had its own ramifications, as the limitations from older formats were carried forward to new ones. The interviewee further noted the need for leapfrogging to better technical formats, whilst making the new model backward compatible so that no data is lost.

All interviewees, except for one who thought that it is theoretically sound but technically complex, saw RDF as the potential metadata format that could handle machine process-able and reconfigurable metadata. They noted that RDF provides better granularity than MARC or other formats currently in use in libraries. Scalable metadata creation that begins with making simple assertions was preferred by interviewees to one that relied on the traditional approach of creating complete metadata records using a document-based record structure. The importance of linking to various metadata data sets and re-using (sharing) metadata from other libraries was considered, by interviewees, as an important component for improving current metadata creation and usage in libraries. However, commercial software vendors are required to base their system design and development around library standards, formats and principles (Study2-Interviewee7).

10.3. Metadata Linking Technologies

Interviewees expressed their familiarity with the technologies associated with Linked Data, whilst many alluded to the complexities of the underlying formats such as RDF, OWL and SPARQL. Development of Linked Data in libraries was considered very slow. Whilst there is recognition among the library community that these technologies can be useful for representing library metadata, so that it can be usable in the web environment, it was also emphasised that Linked Data should not be used for the new technology's or model's sake only. The metadata consultants who worked with library metadata maintained that Linked Data is an appropriate model for scalable library metadata.

The current metadata conversion efforts from contemporary library formats, such as from MARC to RDF, as interviewees observed, were not well thought through and consensually agreed upon. These efforts are mostly experimental and patchy. The interviewees concurred that the linking infrastructure and the potentials for a scalable metadata model present a good case for the adoption of Linked Data in libraries. However, as the interviewees maintained, libraries are not yet in a position to deal with the technical complexities of RDF and the institutional challenges that it would bring about. In regard to this, it is imperative that a proper distinction is made between the conceptual underpinnings and the technical tools required for metadata linking, so that libraries can develop the skill sets in-house or use external expertise. At present, the technical complexities seem to have hindered the adoption of the conceptual foundations of Linked Data.

One interviewee acknowledged the efforts, by national libraries and international organisations such as IFLA, in this direction. In contrast, another interviewee criticised current efforts for their lack of coordination and commitment of adequate resources, remarking

“The work that IFLA is doing to take the standards formats, such as ISBD and UNIMARC, and putting it in RDF is not going to get us where we need to be. It is an interesting exercise, but it isn't really producing [the desired result]. It is much like what we did when we moved from the card to the MARC, where we carried over the same data in the same way, just into a new format. What we need to do is to change the way of thinking about our data. So I kind of understand IFLA doing this, but by doing that I just don't think that moves us very far forward” (Study2-Interviewee7).

Making the distinction between the conceptual and technical sides of Linked Data is crucial. In this regard, it is argued, it is important to delineate contemporary library principles and models such as RDA and FRBR vis-à-vis library record format such as MARC. It is suggested that libraries leapfrog from current formats to RDF and OWL. However, it is mandatory to re-conceptualize the underlying principles embedded in RDA and FRBR, in order to make these library principles compatible with Linked Data. One interviewee seems to contend that “The question again is whether by doing so we will make enough change. Because RDA, after they started developing it, people were complaining and saying that it was too different, and so, they added back some of the practises, such as alphabetical headings.”

As the data analysis showed, in principle, Linked Data is a viable opportunity for libraries to link their metadata with external data providers, which in turn would make library metadata connected to external information sources. However, libraries are not easily going to abandon their principles, which they have developed over hundreds of years. Linked Data principles however do not require libraries to abandon their existing principles and legacy databases. Instead, the question for Linked Data at present is to incorporate four major principles, which would enable libraries to link to external data sources on a global scale. Linked Data principles start from the most fundamental component of Linked Data, i.e. the use of globally unique URI as names for distinctively denoting such things as information objects, people, places, and events (Berners-Lee, 2009). Other essential principles include an ability to de-reference a given URI, using Hypertext Transfer Protocol (HTTP), in order to retrieve relevant information as to what this particular URI refers to; use of data formats, such as RDF/XML, use of vocabulary definition languages, such as RDFS and OWL, and use of query language known as SPARQL; and last but not least is the incorporation of links within and outside the data sets, thereby enriching the data and providing contextual significance (Berners-Lee, 2009).

The interviewees expressed their preferences for the features of links that they would like to see in library metadata:

“The problem of the web at the moment is that the link basically says ‘See also’, it doesn’t say anything about the authority of the linker, it doesn’t say anything about why they are linking, whereas by using some of the standards of Linked Data, you can make a link, you can describe your authority, you can describe the form of the comment you are making, and there is plenty of scope for disagreement, plenty of scope for interpretation. How a system displays that to the end user is an open question. But all that should be encoded in the description, it is all there” (Study2-Interviewee12).

The interviewees tended to gravitate towards particular technologies and formats, such as XML, RDF, Linked Data, OWL and URI, in their discussion (Figure 10.2). They described in detail the features and benefits of these technologies. One interviewee compared Linked Data with prior technologies such as relational database systems. The interviewee noted,

“The power of Linked data is the link. It is the fact that an assertion or an object described in one place, by one person, is linked over the web to things said elsewhere. A RDBMS is a closed box. It is a very powerful closed box, but it is very difficult for a relational database to say something about an object and to know it is the same object described by another person in another database” (Study2-Interviewee12).

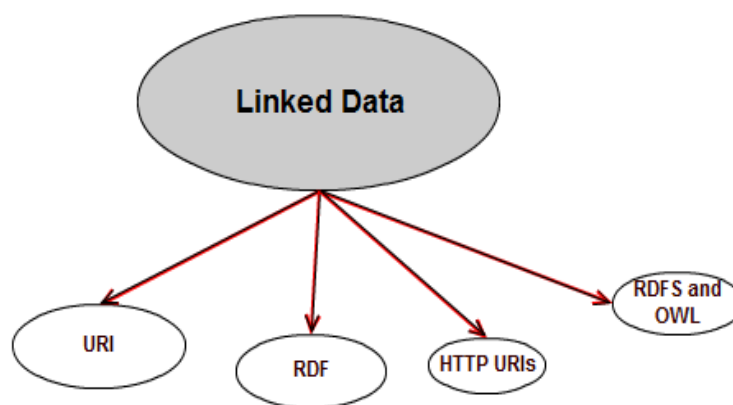


Figure 10.2: Linked Data Technologies

The use of universal URIs, RDF, RDFS and OWL are the main elements of Linked Data technologies. Using Linked Data models, metadata can be represented in simple statements of subject-predicate-object. For example, a metadata statement is fractured in to three (subject-predicate-object) constituent elements such as “*George Orwell* → *isAuthorOf* → *Animal Farm*”.

Linked Data, interviewees pointed out, can help to break metadata records into a series of atomic metadata statements, which then can help to repurpose and recompose to describe information objects. Linked library metadata, it is indicated, opens opportunities to be mixed with other metadata sources outside the remit of libraries. For this to happen, it is indicated, current metadata approaches should be re-oriented so that they support linking across multiple metadata databases.

All the four metadata consultants interviewed (Study2-Interviewee7; Study2-Interviewee9; Study2-Interviewee11; and Study2-Interviewee12) considered Linked Data as an important model to obviate some of the metadata challenges associated with contemporary metadata approaches of creating metadata records as human-readable documents. Linked Data promotes the creation of linking incrementally. The incremental link creation can be made not only by one organisation but the new structure allows other organisations to cross-link to the metadata maintained by others.

Current implementations of Linked Data, as mentioned by interviewees, include the British National Bibliography (BNB) (which is published using RDF), OCLC’s Virtual International Authority File (VIAF), LCSH, Lexvo, GeoNames, MARC country and language, Dewey.info, RDF Book Mashup. Interviewees indicated that current developments in using Linked Data for libraries are in its infancy. One interviewee described the current status by stating the problems of absence of agreement on persistent identifiers resulted in weak (limited) connections between data stores. The interviewee described the development of Linked Data for library metadata as very slow (Study2-Interviewee7).

10.4. Re-conceptualising Library Metadata

In the process of the Theoretical Coding of data analysis, a purposeful decision was made not to dwell on each of these technologies and formats, but instead to put emphasis on the abstraction of the specific technologies to categories that can better represent the interviewees concerns and preferences (see Figure 10.3).

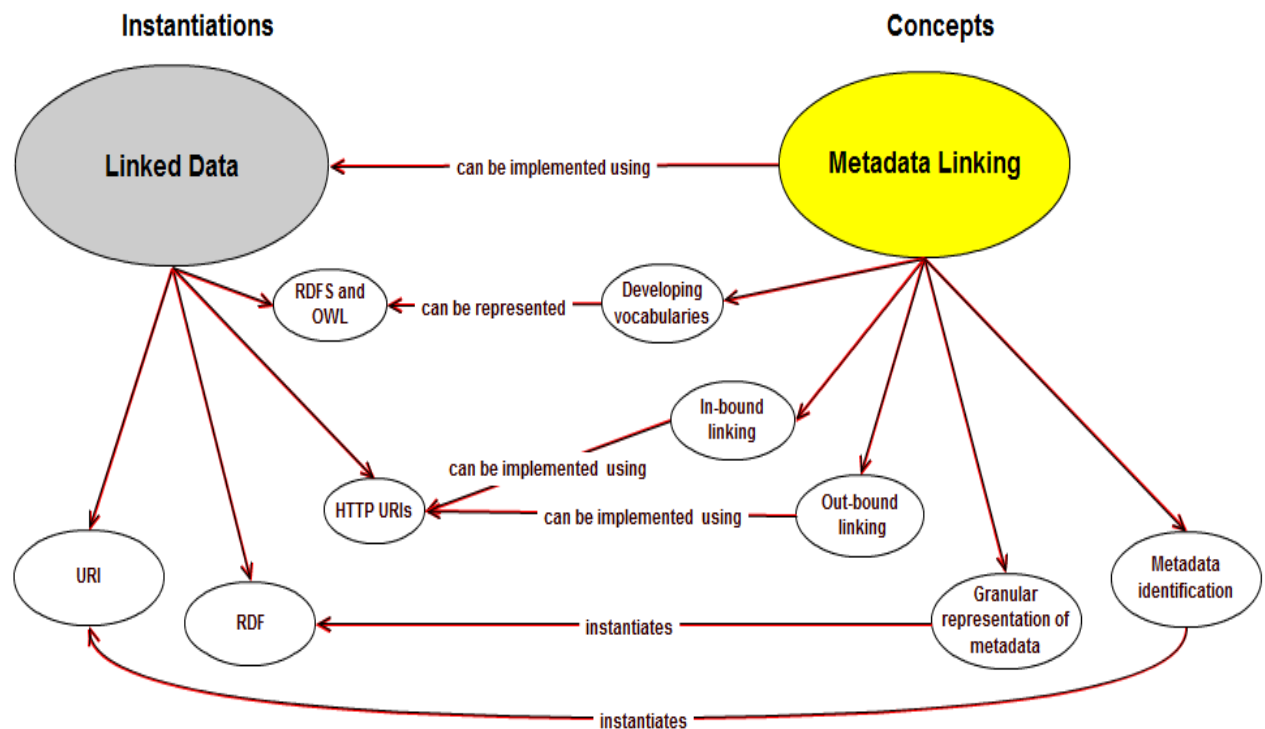


Figure 10.3: Concepts related to Metadata Linking

The technologies are thus used as examples in the memo writing process (see Figure 10.4)

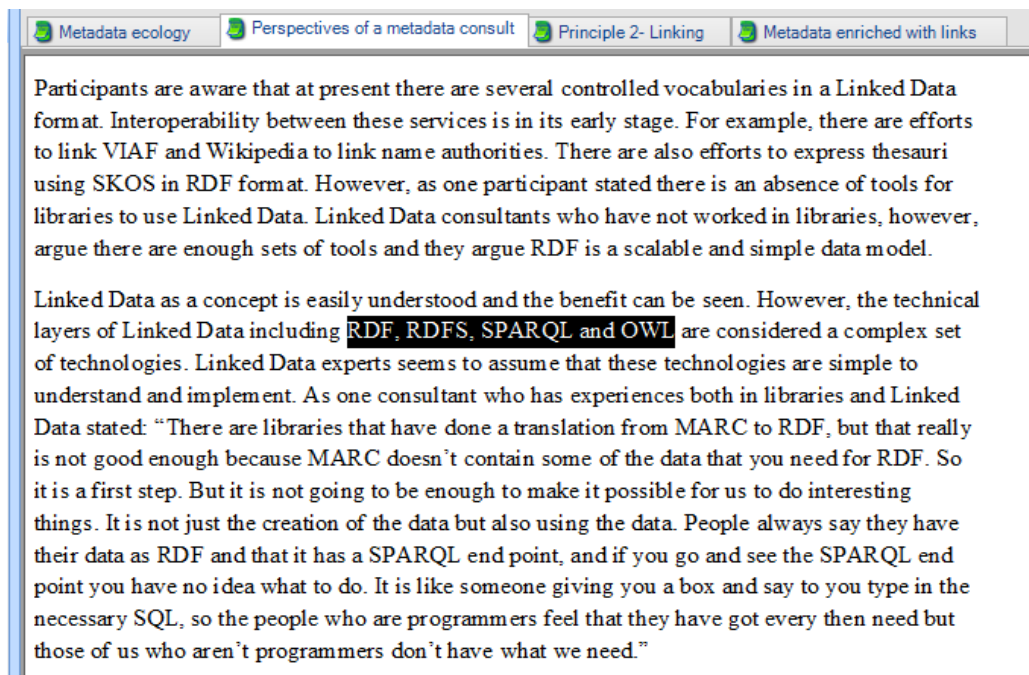


Figure 10.4: Memo on Metadata Linking Technologies

This was also found useful as such technologies are apt to easily become obsolete, potentially affecting the relevancy of the analyses and discussion. Furthermore, it should be borne in mind that not all the interviewees agreed on the relative importance of particular technologies for their specific situations.

The data analysis indicated interviewees' preferences for the creation and representation of metadata that could easily be linked to both internal and external sources. To that end, they underscored the importance of machine process-able, aggregate-able and re-combinable metadata, wherein each metadata value and its associated property (metadata element/field) is uniquely identifiable. For instance, one interviewee dwelt on the importance of decomposing (or atomizing) what had previously been a document-like metadata record structure into a whole series of statements, which could provide maximum flexibility. The interviewee noted that the individual statements can then be joined with other statements, thus providing a more complete description of information objects (Study2-Interviewee7). This revealed the need for a re-conceptualisation of current metadata formats. Figure 10.5 illustrates metadata that is cross-linked (compare this with Figure 10.1).

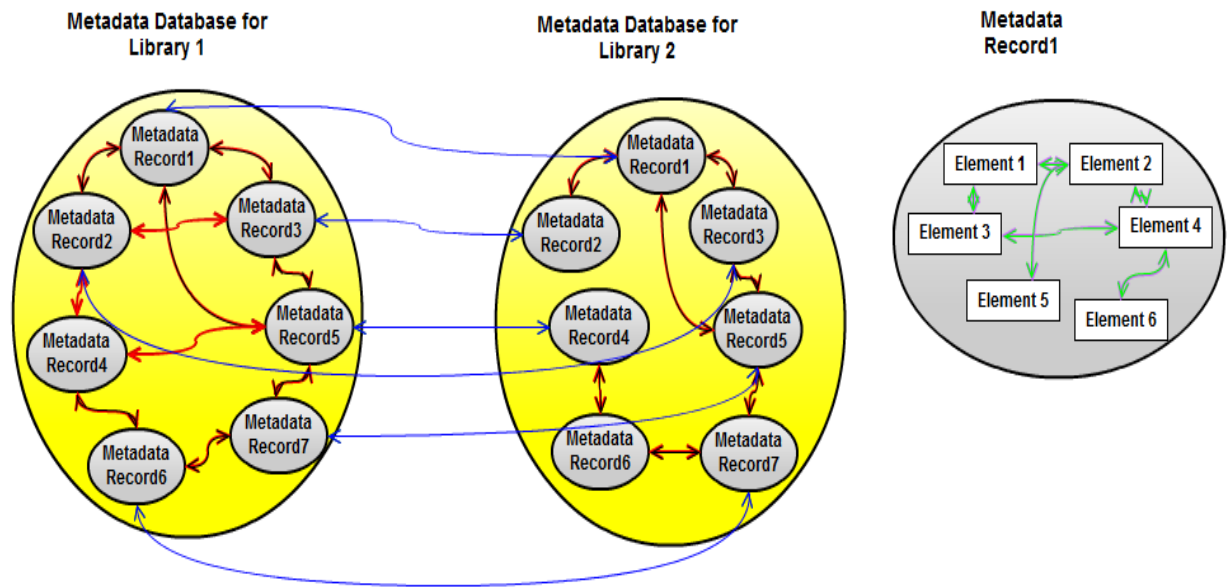


Figure 10.5: Metadata that is Linked

Library agencies at present have been deliberating on finding sustainable solutions to long standing problems of metadata technologies, especially in relation to making library metadata easily interoperable through links. As the interviewees noted, there have been attempts to translate MARC records into MARC/XML, and recently MARC/XML into RDF. At present, as one interviewee observed, some of the world's biggest library institutions, including the Library of Congress, the British Library, the national libraries of France and Germany and Europeana are playing active roles in this arena (Study2-Interviewee4). The re-conceptualisation of metadata representation and linking has several implications on the current use of metadata.

10.4.1. Breaking Metadata Values into Granular Metadata Statements

Links scale if the metadata about an information object is described at a granular or atomic level. As one interviewee (Study2-Interviewee11) noted granular description allows metadata to be mixed and recombined. The same interviewee underscored that greater metadata utility emerges from creating simple metadata statements that scale; each statement can then be dissected and linked to another metadata set based on need. Interviewees emphasised the importance of structuring metadata in such a way that it can be linked not only within an institutional or system's boundary but also in the wider web context.

With such a scalable design of metadata linking, as envisaged by interviewees, an assertion statement can be uniquely identified, and, thus, can be contextualised with other metadata assertions elsewhere (Study2-Interviewee12). The interviewee further elaborated that, by using the same identifier, two or more metadata statements can become linked to each other. The same interviewee, however, noted the limitations of the current Web technologies to develop scalable linking at the data level. The interviewee maintained that Linked Data offers better opportunities for providing solutions to current challenges.

Interviewees (such as Study2-Interviewee6 and Study2-Interviewee11) stated, metadata linking would enable libraries to create cross-links between metadata so that metadata about authors, publishers, subject-vocabularies, and identifiers can be re-used, thus avoiding duplication of effort (see 7.6.3). As the data analysis showed, in order to allow metadata linking at the atomic level (single most element such as a keyword, or name), the underlying metadata structure plays a significant role. As discussed in 7.4.1, metadata structure is one of the Categories which metadata standardisation can provide, albeit current library formats such as MARC are limited in their structure. Metadata that is granularly structured facilitates machine process-ability. Library metadata such as that embedded in the traditional record format is stored and exchanged by machines but it predominantly tends to remain document-centric where human users must painstakingly check and use it as opposed to machine process-able metadata that can automatically be re-used and re-purposed.

10.4.2. Unique Metadata Identifiers

The interviewees discussed the issue of uniquely identifying information objects, variously citing the examples of specific instance techniques such as the use of ISBN, URI and classification numbers. The importance of identification mechanisms that scale and interoperate is considered an important feature of linking library metadata. The role of namespaces to uniquely and globally identify metadata sets and values was indicated (Study2-Interviewee4; Study2-Interviewee9). There are currently several international and institutional efforts to develop namespaces for unique identification. However, it was indicated that best practises on the use of global identifiers for metadata is yet to emerge. As one interviewee described the use of an identifier allows recombining various metadata elements and values from disparate sources (Study2-Interviewee4). One interviewee described the current status and future prospects of globally identifying metadata as: “If you like to link back to information about a book in the library world, today that is a very difficult thing to do because there is no universally agreed upon identifier for that book” (Study2-Interviewee11).

One of the challenges of identity in libraries has been that the usage mechanisms have been predominately used within libraries themselves, and not by non-library service providers, thus making it a challenge to cross-link disparate resources. Current efforts at finding mechanisms for identification (such as the IFLA Namespaces Group, the British Library Data Model, and the Europeana Data Model) focus on standards-based metadata, there is little effort expended to develop identification mechanisms for socially-constructed metadata.

In order for it to be re-usable and shareable, metadata should not only be uniquely identified but the identification mechanism should also be persistent. For this to happen, there should be a concerted co-ordinated effort to develop global identification schemes for metadata elements (properties) and their corresponding values. Due to the apparent enormity of the effort to uniquely identify each metadata value, the efforts should be undertaken both at decentralised and centralised levels. Whilst there are efforts at devising identification mechanisms for standardised metadata, such efforts are not developed to include socially-constructed metadata.

10.4.3. Uniquely Identifying Socially-Constructed Metadata

The case for the importance of socially-constructed metadata in the light of the goal of metadata enriching (see Chapter Nine) has been made. As the interviewees indicated, it is important that standards-setting agencies and libraries work towards the development of unique identification mechanisms for socially-constructed metadata, so that re-usability and mash-ability of such metadata can be realised. Until such identification mechanisms are found, socially-constructed metadata will largely remain constrained to the system in which they were created.

In addition, as the interviewees mentioned, the provenance of metadata (8.4.5.8) can only be made possible if particular user's metadata can be uniquely identified and located, so that the context of its use and who has created it is can be determined. One interviewee noted that "there are many suppliers that are mining text, mining twitter streams and mining status updates on Facebook to identify concepts, meanings, emotions, sentiments, and things like that, but at the present moment all these systems are very closed. But there is no agreed international identifier at the moment" (Study2-Interviewee11). The use of identifiers for social metadata, as the interviewees indicated, is not developed very well. It is thus crucial that unique identification mechanisms are developed for socially-constructed metadata.

In connection with the above, consider tagging. This is the most structured type of socially-constructed metadata and it has three major components: user, tag and information object. The user is a person who tags an information object. Each of the components can be considered as a class and instances can be identified by a Uniform Resource Identifier (URI) scheme. A URI is a global unique identifier which forms a fundamental part of Linked Data (Allemnag & Hendler, 2008). For example, the following example is taken from LibraryThing.com (see Figure 10.6).

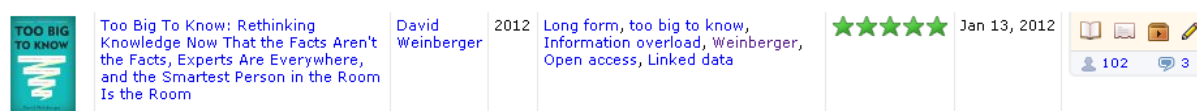


Figure 10.6: Example of Socially-Constructed Metadata (Source LibraryThing.com)

Interviews saw benefits of linking library metadata with outside sources such as Amazon. In addition, library metadata can be linked to other library metadata thus avoiding duplication of effort. As one interviewee noted "if the books are not available on the shelf, there is a link to take you off where you would be able to purchase the book" (Study2-Interviewee2). Such links may also be integrated to sites such as LibraryThing where users are offered the possibilities of looking for information about a particular book, whether to purchase it or read reviews about it. Linking could enrich the OPAC/discovery interface with more relevant information from outside sources. For example, some users may want to check the biography of the author of a book, which is not recorded in the library metadata but exists on outside sources. Another example, mentioned by interviewees, is book covers which can be copied from external providers, such as Amazon, rather than the library trying to recreate them in-house. Book jackets offer users a visual aid to identify books on shelves and electronic lists

(Figure 10.7). Please note that even though book jackets can be considered a priori metadata (for example as used in Amazon) but as exemplified in LibraryThing, they can be contributed by users, thus making it a constituent of post-hoc metadata.

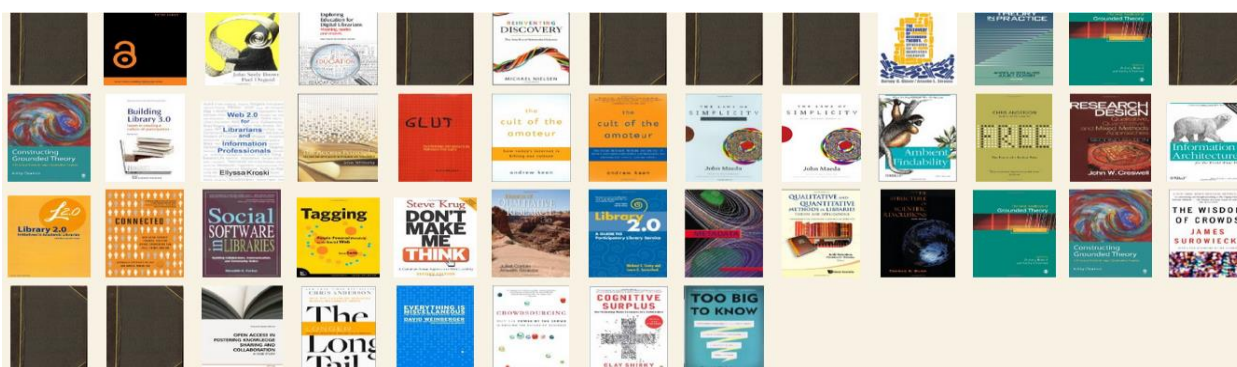


Figure 10.7: Example of Book Covers used in Navigation

In a socially-constructed metadata approach, users can contribute not only tags, ratings and reviews but also images of book covers.

Hence, socially-constructed metadata can be uniquely identified and linked so that it can be mixed, matched and re-used. One interviewee mentioned the possibility of sharing tags if it is linked and made available to other users (Study2-Interviewee4). Linking metadata and making it available for others to re-use, interviewees noted, adds its return on investment as other libraries and applications may use it for various purposes. One interviewee (Study2-Interviewee6) suggested the possible use of library metadata to generate citations automatically.

Once metadata is uniquely identified and made accessible for cross-linking, there are several benefits to be gained. For example, one interviewee noted, “I think there is just so much to be done with these metadata and joining it up so that it is not all of us sitting in our little silos of data [which may include the ability to connect with other universities”. They also saw benefits in a nation-wide integrated catalogue where one could check what books other universities have and for a user to go and read these books.

10.5. Benefits of Metadata Linking

Links are a powerful means to improve navigation and browsing experiences by making it easy for users to jump from one information object to another. As one interviewee noted, links provide mechanisms to link from one source to another, to say “this movie is adapted from this book, or this book is based on this other book” (Study2-Interviewee7). Links, if consistently designed and implemented, provide scalable means to enrich metadata as different people can create connections between information objects. In other words, the critical mass of links is beneficial to enrich metadata with contextual relations. That is, “people will get interested, those people who have read a lot on a topic, will go there and make a couple of connections, somebody else will make a couple of other connections, and then you reach a kind of a critical mass of all these connections” (Study2-Interviewee7). Metadata values that are not linked indicated to low visibility by either machines or human

users (see Chapter 7.6.2). One interviewee noted that present library metadata has low visibility that affects the library collection's discoverability (Study2-Interviewee7). More links, the interviewee noted, is the solution to these challenges.

The benefit of linking is that library resources become more usable. One interviewee indicated that "people will be able to find library resources whilst they are doing whatever kind of information seeking on the web" (Study2-Interviewee7). Attributing Vannevar Bush's idea of the memex, one interviewee cited the importance of the notion of associative trails or links between documents. Associative trails, the interviewee indicated, are one of the great benefits of rich links, whilst also noting that it is also how users get lost navigating and discovering through the links.

"Because, we want to know how things are connected to one another, one of the ways we learn is by making those connections. So if I am to describe what the inside of an eyeball is like, I would say it is something like jelly, if you've encountered jelly before you know what I am talking about. If we have digital object of whatever kind, not only that document has some kind of creator, somebody must have made it, somewhere, somehow, but documents themselves belong to a whole web of other documents. No document exists by itself, not only is it linked to other documents, but to other ideas" (Study2-Interviewee8).

As one interviewee described one of the advantages of metadata linking is that it can be enriched with more links in an incremental manner. In addition, linking also means the library should not necessarily create all the metadata description in-house, instead links can be created to external metadata descriptions. For example, a link can be made between a metadata value in an OPAC display to a socially-constructed tag, rating or review in LibraryThing.com.

Interviewees also stated the benefits of metadata interoperability to facilitate metadata aggregation and metadata cross-linking which in turn would improve cross-searching of information systems.

In contrast to current efforts that focus on incorporating formal and authoritative metadata, most interviewees accentuated the importance of incorporating metadata descriptions created by users. To this end, it is indicated that as metadata is aimed to benefit users, it should not be 'jargonised', which might be difficult for users to understand. One interviewee said, "I don't need to have a library science degree to understand the metadata" (Study2-Interviewee11). In addition, one interviewee indicated that the multilingualism aspect of metadata should be considered. Metadata multilingualism was in particular thought to be useful to represent controlled vocabularies such as subject headings (Study2-Interviewee11). It is important to note here that the idea of linking metadata values is mainly dependent on the design of metadata identifiers, structure and granularity. Including users, linking to non-library sources, multilingual metadata, harnessing local and international metadata descriptions all seem to indicate to an underlying requirement for metadata diversity (9.2.1).

Interviewees noted the particular relevance of linking library metadata with outside sources such as LibraryThing. One interviewee pointed out that, “if the books are not available on the shelf, there is a link to take you off where you would be able to purchase the book”. Such links may also be integrated to sites such as LibraryThing where users are offered the possibilities of looking for information about a particular book whether that is to purchase it or read reviews about it. Linking to external sources such as Amazon or Google Books, as one interviewee (Study2-Interviewee5) indicated, benefits users to read tables of contents and free chapters so that they can determine the relevance of particular information objects for their study or research, thus indicating the importance of purposeful linking to such sources. Users may also want to buy information objects rather than borrowing them from the library. One interviewee saw no problem with making such processes easier for library users.

Links add value to the experiences of users by providing contextual and relevant connections between resources, which traditionally seems to have appeared disconnected. Relationship between concepts in a search result can be linked to subject headings which have structured semantic links between broader, narrower, synonym and related terms. One interviewee (Study2-Interviewee4) emphasised that sharing of metadata, crowd sourced tags (8.4.1.5) and the idea of a network effect (8.4.5.5) for metadata can only be achieved through the use of effective linking mechanisms. As one interviewee (Study2-Interviewee6) noted, library metadata should be semantically described and granularity and machine process-ability should be sought so as to utilise such metadata for various purposes other than library cataloguing. One such use mentioned by the interviewee was citation analysis where similar names of authors and their contributions can be disambiguated so as to also ease the process of citation impact analysis.

10.6. Conclusions

The interview data was analysed in relation to the interviewees’ perspectives on the implications and benefits of linking library metadata with external data sources. In order to optimally benefit from metadata enriching, libraries should aim to harness both in-bound and out-bound links (see Figure 10.8).

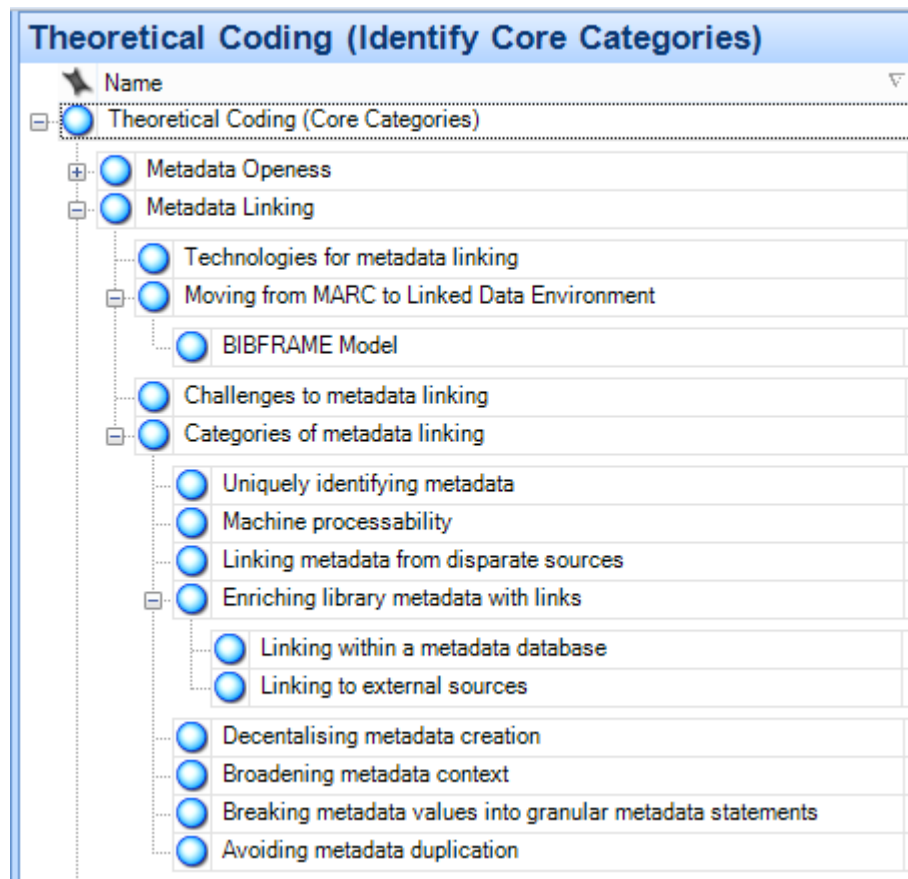


Figure 10.8: The Core Category of Metadata Linking

Effective linking requires a metadata design that takes into account the unique identification of metadata elements and corresponding values. Unique identifiers that persist facilitate metadata sharing and interoperability. Since metadata enriching is provided through the combined use of standards-based (a priori) and socially-constructed (post-hoc) metadata, it is of paramount importance that both strands of metadata are uniquely identified and maintained. Metadata that is designed with granular structure and metadata that is enriched both with in-bound and out-bound links helps to minimise or avoid duplication of metadata records with-in and among institutions. It brings disparate metadata sources together. It widens opportunities for collaboration and opens vistas for novel uses of metadata that may not have been envisaged by the original metadata creators.

The findings, with regard to the importance of linking metadata, are consistent with emerging web developments, especially with that of Linked Data and the Semantic Web. Whilst the standards-based legacy metadata resulted in read-only (Web 1.0) metadata where users have a passive role of merely searching or browsing through library collections using the library catalogue, socially-constructed metadata approaches result in read-write metadata (Web 2.0) where users are pro-active co-creators as they can participate in the creation and management of metadata. However, in what would be considered as a web 3.0 approach, as the grounded theory data analysis in this research showed, the strengths of both standards-based and socially-constructed metadata approaches can be brought together in a mixed metadata approach, thus metadata that is enriched through socially-constructed metadata can be

structured, aggregated and analysed providing a collective metadata intelligence. For all these to take effect, the power of links and relationships is obviously indispensable.

Chapter Eleven: The Core Category of Metadata Openness

11.1. Chapter Overview

A procedure similar to that employed in the previous two chapters was followed, whereby the Theoretical Coding of interview data led to the emergence of the Core Category of metadata openness, which forms the topic of this chapter. The analysis showed that standards-based metadata that is enriched through socially-constructed metadata can be effectively linked to metadata that resides in both internal and external information sources only if metadata openness is embraced as a mechanism that facilitates sharing and re-use, thus further benefiting all stakeholders. The interviewees pointed out the importance of making bibliographic data openly available for sharing, rather than guarding (hoarding) it. The main aims of metadata openness, as the analysis showed, are ensuring institutional transparency and accountability; increasing returns on investment on metadata; metadata sharing and re-usability and mashing-up metadata from disparate sources.

To begin with, it is important to make an important distinction between the terms free metadata and open metadata. Whilst free metadata directly refers to the waiving of the cost, which is associated with the creation and organisation of metadata, the connotation of openness refers to the freedom to access the actual metadata for re-using, mixing, mashing and re-purposing (Tapscott & Williams, 2010). Put another way, metadata openness is a conceptual/theoretical concern whereas the cost of metadata is a practical/technical issue. The focus in this chapter is on metadata openness.

In addition, three types of metadata openness were identified: open metadata, open metadata formats and open-source library management systems software. Whilst open-source library management systems software is outside the remit of this research, the first two will be dealt with here. The analysis indicated the importance of promoting metadata openness through the institution of open licenses that, not only permit cost-free usage, but also encourage adaptation and even commercial-use, so as to optimally utilise metadata not only for cataloguing but also for other purposes within and outside the library environment.

11.2. The Core Category of Metadata Openness

As Focused Coding of interview transcripts that pertain to the perspectives of LIS professionals on making library metadata openly available to external users resulted in a number of categories, which can be considered as categories of metadata openness, and are discussed in the subsequent sections. These categories are discussed in the following sections.

11.2.1. Improving Institutional Transparency and Accountability

As part of a wider government initiative aimed at transparency and accountability, public-funded institutions, such as libraries, are being required to make their bibliographic data openly available (in a form that is accessible and usable). For example, as two of the interviewees stated, the British Library has made the British National Bibliography (BNB)

openly available for the first time since the 1950s. The interviewees noted that the quest for metadata openness is part of an emerging global trend towards open data and open access to scholarly publications. Libraries, the interviewees noted, are always about free access to information, however, bibliographic data has been predominately kept in a closed access environment. It is important to note the distinction between metadata being free vis-à-vis it being open, indicating the broader implications of the latter. Metadata that can be freely available can still be locked in proprietary formats, making re-usability and adaptation all but impossible. Hence, as the interviewees suggested, institutions should exert effort to use transparent and open metadata formats.

As the interviewees noted, libraries are cognisant of the on-going open data movements across the world, and hence they have, quite recently, been actively engaged in regional and international discussions (e.g. conferences, workshops and discussion fora) on the topic. As a result, librarians seem to agree that metadata openness is an important concern that needs to be addressed. One interviewee (Study2-Interviewee12) believed that “there is no reason for the basic bibliographic data to be locked away”, whilst admitting that metadata openness should be dealt with at a national library level, so as not to burden every local authority library with concerns associated with making its bibliographic data openly available, since, as the interviewee noted, such efforts and risks could more effectively and efficiently be handled in a centralised, rather than a decentralised, manner. This is because, at present, local libraries purchase a bulk of their bibliographic metadata records from vendors such as OCLC or from their own national library. As the interviewees attested, due to resource constraints, most academic libraries exert little effort to enhance and customise the metadata obtained from external sources. For instance, libraries could not easily add a new metadata field that addresses their local requirements. Thus, it was suggested that the efforts of metadata openness should not itself be duplicated, thus obviating the importance of collaboration and coordination.

Another avenue for openness, as one interviewee indicated, is the ability for institutions and individuals to develop new business models where value can be accrued from the merging of disparate metadata in novel and interesting ways (Study2-Interviewee4). It was also indicated that current global technological and business trends are beginning to prove the importance of such openness.

Libraries are thus beginning to realise the strategic importance of metadata openness, not only for alerting their users to the list of available information objects but also to enable other institutions, including libraries and system developers, to re-use it. However, it is indicated that, the shift from traditional metadata silos to openness requires a concerted effort as it has resource and policy implications.

As the interviewees pointed out, it is only upon making library metadata openly accessible for re-use that its full benefit can be realised. Put another way, it is not easy to predetermine how one institution’s data can be re-purposed, along with other data, so as to benefit users. To implement metadata openness is therefore considered an important institutional challenge,

which has wider implications on the overall aspirations to metadata enriching (see Chapter Nine) and metadata linking (see Chapter Ten).

11.2.2. Return on Investment

In addition to contributing to ensuring transparency and accountability, metadata openness is believed to bring additional returns on investment. Since the efforts to create and maintain metadata incur costs, as the interviewees noted, opening the metadata for a wider use may bring economic advantages. One interviewee (Study2-Interviewee4) reiterated a question that they said they have been posing for themselves: “Can people begin to develop new business models using metadata from a variety of different sources that the originating organisations may not be able to devise for their own institution?” The interviewee added that the answers to this question can only be seen if metadata openness is implemented and widely adopted. The major driving force towards metadata openness, as the interviewees noted, is the belief that sharing rather than locking institutional data would bring overall economic and societal benefits. One interviewee (Study2-Interviewee11) noted that library metadata should be liberated from its silos in which they thought that the silos are locked through the use of library-centric data formats such as MARC. In addition, institutional policies require that library metadata be protected through intellectual property rights.

“You know, we’ve been capturing, curating and archiving data for many years, now [we are seeing ourselves] moving to seeing that data as a raw material to build new things, manufacture new things, it is very analogous to the industrial revolution where people were aware of the value of storing iron and coal for many centuries. We have to learn how to take our raw material of data and to use that data”.

One of the motivating forces for metadata openness is the ability to re-mix and mash-up data from disparate sources. Freely and openly available databases such as Wikimedia’s DBPedia database, and the ability to re-use such data through application programming interfaces (APIs), it is emphasised, provides an essential impetus for the call towards metadata openness. As metadata is shared across multiple institutions and applications.

11.2.3. Metadata Sharing and Re-usability

Metadata sharing and re-usability are important benefits of metadata openness. Metadata sharing and re-usability can be effected either through purchase or via free access to metadata. As interviewees indicated, if the route to metadata openness is followed, incremental metadata creation and enhancement can be made possible, through linking to metadata that is stored outside the remit of the library (Chapter Ten). As most interviewees observed, currently extant bibliographic data, which contains millions of records expressed in atomic statements, could be linked to subject headings expressed in different languages.

One interviewee (Study2-Interviewee4) described the case of a project in which their library redesigned their metadata system in such a way that it was linked to external metadata sources such as VIAF, LCSH, Lexvo, GeoNames, MARC country and language, Dewey.info and RDF Book Mashup. As a result of the success of this endeavour, the interviewee added,

their library is currently planning to link their library data to external sources, such as DBPedia, which is openly accessible as Linked Data. It is believed that interesting relationships may be made between the data elements and useful patterns may emerge, adding value to users (Study2-Interviewee7).

As one interviewee (Study2-Interviewee4) remarked, metadata openness supports automatic aggregation, from which novel ways of metadata analysis could emerge. As the interviewee pointed out, subject terms from one source can be cross-mapped to its equivalent term in another metadata source. It was further noted that similar cross-mappings could be made between socially-constructed metadata residing in different applications. In addition metadata aggregation can be implemented through the use of links to book covers, reviews of books, author biographies, and related and contrary works residing in disparate data sources

Interviewees believed that libraries would benefit from embracing openness of their data so that linking among the various data sources is possible. One interviewee noted that openness enables linking between library metadata to social metadata, such as LibraryThing.com. The interviewee added that “I think there is just so much more yet to be done with all these existing metadata and joining it up so that it is not all of it sitting in our little silos of data” (Study2-Interviewee2). Another interviewee (Study2-Interviewee1) noted that openness would foster efficient use of resources, such as between libraries within a region or country. The interviewee emphasised the importance of openness for cooperation and sharing, so that the library not only serves the users within its remit but also the community at large. One interviewee supplemented the importance of connecting information objects across universities.

“How brilliant would it be if, in effect, there was a UK universities catalogue where you could search? Let us say, ‘I am going to Manchester for the holidays, what books do they have there?’ Yes, you could now go and search each individual catalogue, but wouldn’t it be nice if you could search just one catalogue? I just think there is much more that we have not begun to explore yet, in terms of shared data, especially joining into other things like LibraryThing as a kind of social phenomenon, Amazon as a commercial entity, that kind of thing” (Study2-Interviewee2).

One interviewee (Study2-Interviewee11) added that library metadata, including socially-constructed metadata, should be made openly available so that for example tags, book reviews and recommendations can be mined. The sharing of tags (pulling tags from disparate applications) and tag crowd-sourcing are considered important in the light of metadata openness (Study2-Interviewee4). Another interviewee (Study2-Interviewee2) noted that the future of metadata, as they saw it, would be increasingly open, creating possibilities for re-use such that it optimises the experience of users, as the latter are interested in accessing, rather than being concerned with institutional and technical intricacies that lie behind the scenes. Thus, openness benefits users tremendously since it broadens the sphere of metadata access, thereby increasing the amount of content that they get. In other words, users, as the interviewee noted, are not interested in the location of the information source, as long as they get access to what they need, at the right time.

11.2.4. Opening Enriched Metadata

As the interviewees noted, the value proposition for metadata openness does make sense, as metadata can add value because it incorporates unique data. Unique metadata, however, can only be made available if libraries enhance the metadata created by external sources, perhaps by involving users in the metadata creation process through socially-constructed metadata approaches.

Interviewees indicated that if libraries would make a case to users that the metadata they created would benefit other users, it would perhaps serve as an important motivation for socially-constructed metadata (see also 8.5). In such cases libraries have an additional responsibility to make sure that metadata is effectively aggregated and organised.

“It depends on the wider picture. In principle, I am very attracted by the idea that more and more data should be made openly available. But again, it depends very much on context and what you are hoping to get people to want to contribute” (Study2-Interviewee13).

Opening up socially-constructed metadata has privacy and intellectual property implications. As the interviewee above (Study2-Interviewee13) noted, it is likely that different users would have different views on the metadata that they have created. “Some [users] may be reluctant to contribute to an open space, but very willing to contribute to a closed one. A kind of private community of experts in a particular field might want their contributions to be accessible to that particular community alone. On the other hand, if you are looking at the sort of crowd sourcing data projects, such as transcription projects, you have a large number of people contributing to the project they may expect that the contributions become openly available, benefiting others”.

11.2.5. Improved User Experiences

Open metadata can augment the user experience through its ability to facilitate the extension of relevant metadata beyond what is available within a particular library, to what is available at external sources. One interviewee noted that such openness would allow users to explore for and discover new information objects (Study2-Interviewee2). The interviewee maintained that metadata openness would also enable libraries to link to external sources that are outside their remit, thus facilitating exploration and discovery of information objects by users.

One interviewee pointed out the importance of sharing metadata across libraries, especially among those which are publicly funded in a given country, thus obviating barriers to access. The interviewee indicated that complex authentication mechanisms at present make openness an intractable challenge, but nonetheless, they went on to note, the need for improving the experience of users by removing unnecessary barriers, especially if the resources supporting this access have already been paid for nationally.

With regard to socially-constructed metadata, interviewees saw no reason why such metadata should not be made openly available across applications, citing, for example, the benefits of

linking library metadata with Wikipedia so that users would be able to explore more about the biography of authors and other works written by them.

11.3. Degrees of Metadata Openness and Metadata Licensing

Metadata openness is not about making data freely (cost free) available but it indicates the ability to access the data in a form that is accessible and re-usable. This has implications on the kind of licenses associated with the metadata that is made available. Traditionally, data could be shared and re-used, but further adaptations and commercial re-use were largely forbidden. The concept for metadata openness, as the interviewees indicated, should include adaptation (without compromising intellectual integrity) and commercial use. The rationale for such openness, it is indicated, is the assertion that even if such adaptations and commercial use may not directly benefit libraries, it would accrue overall benefits for society as a whole (users). As indicated by interviewees, metadata openness should thus be mandated through licensing schemes and procedures. One of the examples of such licenses is the Creative Commons Zero (CC0) license. As one interviewee noted, this licensing scheme provides a flexible use of data for re-usage, re-mixing/mashing up, adaptation, and even commercial use.

“We tried originally on a non-commercial option, an attribution option, but realistically, we’ve got the voluntary attribution, but we’ve accepted that people, especially those who were pursuing the government’s agenda to try to encourage economic benefits, have to have the commercial option. So we’ve used commercial licensing, which has the added benefit, as far as we are concerned, although it has not been fully tested in the courts, of being able to claim no liability. So if somebody takes our data and does something strange with it, then it’s not our fault that somebody else has done something with it. So that’s been quite a useful thing as well. No licensing system is perfect; there are many ways of freeing data” (Study2-Interviewee4).

The permission for commercial use encourages application developers and commercial companies to leverage the existing data in various ways. For example, as indicated by interviewees, authority data such as lists of authors, publishers, geographic names and overall bibliographic data can be used in novel ways using current technological trends. Thus, the use of very open licenses, the interviewees indicated, should be encouraged, or even mandated.

Currently, as one interviewee identified, there exist a five-star rating system that provides information about the degree of openness of data. The ratings range from very minimal degree of openness where only the cost of metadata is free but still in proprietary (locked) data formats (whereby adaptation and commercial uses are prohibited) to those that are most flexible (where data is not only freely available but also made available in open and machine process-able formats which in turn can be adapted and used for commercial purposes). The interviewee (Study2-Interviewee4) noted that their library aspires to attain a five star-rating status as they plan to embark on the unrestricted metadata openness route for their bibliographic data holdings. Figure 11.1 details this five-star rating scheme, within which publicly-funded organisations are aiming to make their data available on the Web.

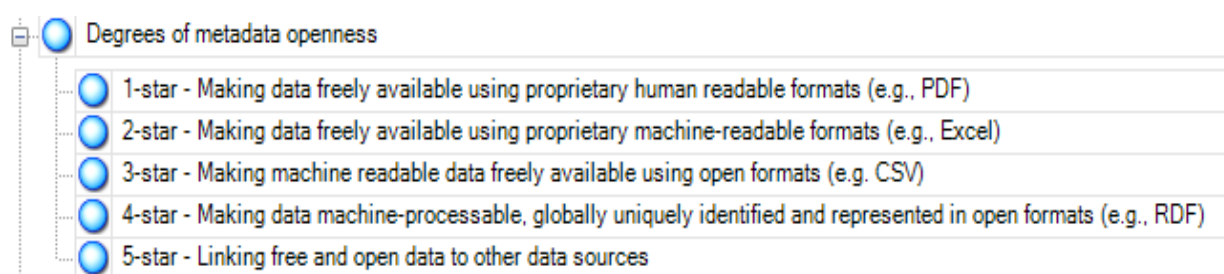


Figure 11.1: Degree of Metadata Openness

A single star rating implies that metadata available for free but it is in a format that is only human-readable. Strictly speaking, metadata that is licensed using a single star is cost-free but not open. Whilst metadata obtained for free has a significant benefit for libraries in terms of saving money, a single rating scheme does not make the data easily re-usable. In connection with this, most MARC data, as the interviewees indicated, could be considered closed since the interviewees argued that a MARC record is a flat file. As one interviewee (Study2-Interviewee4) described, MARC was designed to create self-contained electronic metadata descriptions, which nonetheless cannot be effectively processed by applications through aggregation or mash-ups. Another interviewee (Study2-Interviewee11) compared MARC to a card catalogue that is stored in an electronic format, thus adding to the challenges of re-using the data embedded in it. It is thus safe to assert that the data embedded in MARC could at best garner a single star-rating. As the interviewees remarked, this is not surprising, given the age of MARC which dates back to the 1960s where computer data formats were not as developed as they are today. Even if there is now consensus that metadata openness could benefit libraries, the issue of metadata formats and the significant amount of metadata currently residing in these old formats implies that the goal of attaining a five-star rating is a long way off.

Whilst the increase in the degree of openness is beneficial, the ideals of metadata openness could be optimally utilised if it is not only represented in open formats and is freely available, but in its being linked (see Chapter Ten). Thus, a five-star rating should be the goal if a library aspires to fully embrace openness. A five-star (*****) rating indicates that metadata is already rated as four-star is enriched with links both within the database and also to external metadata (Chapter Nine). In order for metadata openness to fulfil a five-star rating criterion, metadata values should be globally and uniquely identified, represented in open and scalable formats, enriched with links to internal and external data sources (Chapter Ten), and amenable for re-use, mixing and matching with other data sources. Metadata that has a five-star rating encourages re-use and aggregation between data sources, thus data owners should make explicit what licensing scheme and star rating they impose when making their data available. Licensing ensures users explicitly know, beforehand, the nature of data as well as the rights issues that are associated to it. It should be evident that if a metadata database is made available with a five-star rating, it means that it has achieved the highest degree of openness.

As one interviewee (Study2-Interviewee4) stated, the use of Unique Resource Identifiers (URI), RDF and related Linked Data technologies could help attain a five-star rating:.

“We are beginning to look at new ways of expressing that data, especially if we want to get it outside, to increase the utility of the value that we have put in to the descriptive metadata ourselves for the wider community. It’s a return on investment kind of argument. Is the benefit predominantly for the world of libraries or is it for the wider community? If it’s for the wider community, then we have to get our data out there in forms and via channels that are appropriate. To do that, we are looking at a variety of different formats and a variety of different technologies to enable that to happen. Linked Data is one of them, putting the data into RDF is one, particularly in the academic library. When we’re dealing with developers, who have wanted to create for example, apps for iPhone and the like, using bibliographic data and potentially wanting to merge it with other data” (Study2-Interviewee4).

11.4. Conclusions

Metadata openness is one of the Core Categories that emerged during the Theoretical Coding stage of the data collected from LIS professionals (see Figure 11.2).

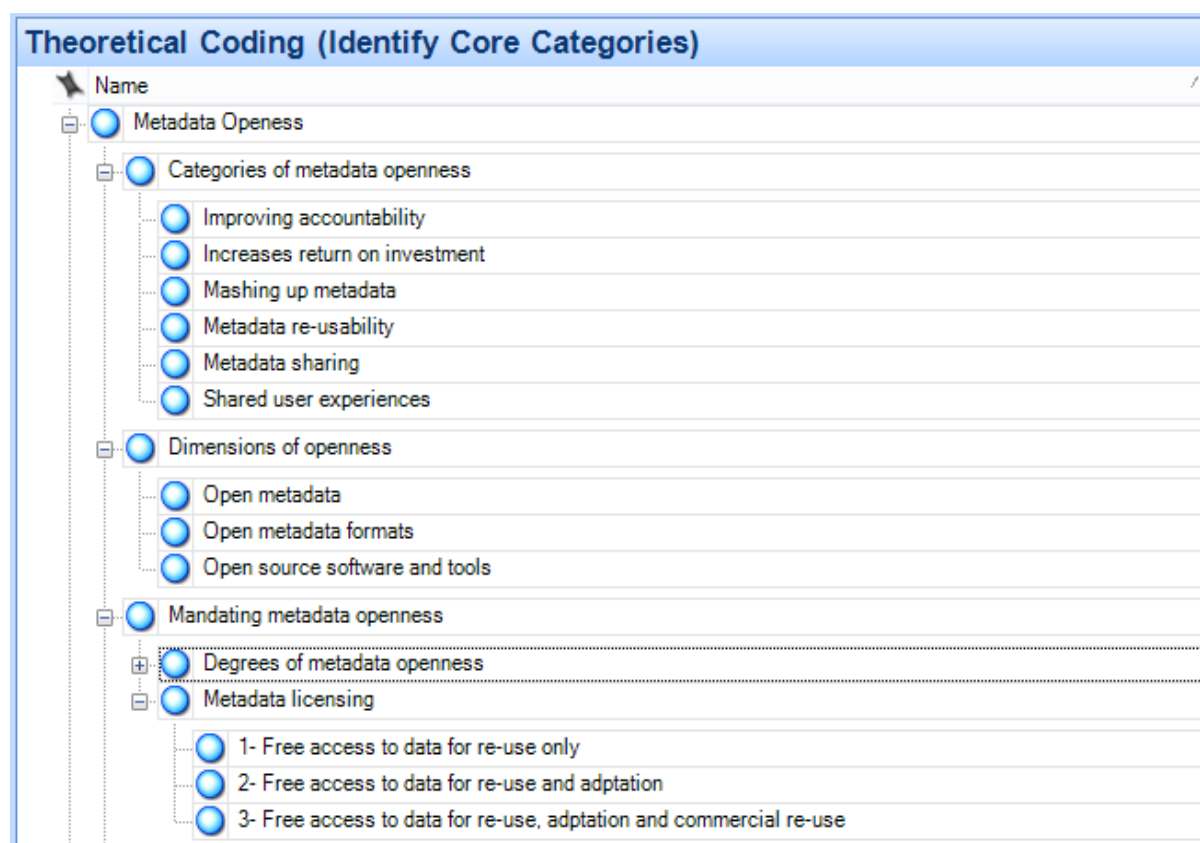


Figure 11.2: The Core Category of Metadata Openness

It was established that, along with quests for open access to scholarly communication, open source software, open and transparent government initiatives, the call for open data has become an emerging requirement for publicly-funded institutions to make their data openly available. Libraries being publicly-funded institutions and library metadata being data in the public domain cannot be exempted from this requirement. The value of standards-based

metadata that is enriched through socially-constructed metadata (metadata enriching see Chapter Nine) increases with the degree of openness (11.3).

The principle of metadata openness is a crucial component of any metadata system, without it libraries cannot implement complete and effective metadata linking (see Chapter Ten). With a growing interest in Linked Data (10.3), the importance of metadata openness has become all the more apparent. As the analyses indicated, metadata openness improves institutional transparency and accountability; increases return on investment and facilitates metadata re-use and sharing, while precluding or minimising data duplication. The analyses indicated the variation (degrees of openness) in order to make library data available on the web, so that users could easily determine the associated permissions as well as the nature of the data in relation to the re-usability of the data formats. Thus, the Core Category of metadata openness emerges as an important conceptual/theoretical theme. Put another way, cost (being free) is a practical issue whereas openness is an intensely theoretical concern which libraries should consider.

Chapter Twelve: The Core Category of Metadata Filtering

12.1. Chapter Overview

A Mixed Metadata Approach is considered an important emerging paradigm where a priori and post-hoc metadata are considered as complementary (Chapter Nine). The mixed metadata paradigm is underpinned by emerging metadata principles identified and developed from the Constructivist Grounded Theory data analysis. These emergent principles included: metadata enriching (Chapter Nine), linking (Chapter Ten) and openness (Chapter Eleven). In connection with this, the interviewees believed that once metadata is enriched, linked and opened, if metadata is presented to users without any mechanism for filtering, the volume of descriptive labels, tags, links, comments, reviews, or ratings, it would confuse users and negatively affect their experience. The Focused coding stage of the grounded theory analysis resulted in the emergence of categories including **providing multiple metadata views** through post-hoc filtering (12.2), **interface simplicity** (12.3), **personalisation** (12.4), **recommendations** (12.5; 12.6), **integrated** and **serendipitous access to information sources** (12.7; 12.8). The Theoretical Coding analysis led to the emergence of the Core Category, i.e. metadata filtering.

12.2. Providing Multiple Metadata Views through Post-Hoc Metadata Filtering

At present OPACs provide a single global view to all users. Thus different users searching the same keyword or following the same navigation path find the same result sets. The interviewees considered this a serious limitation. The single metadata view is not perhaps very problematic, when the size of current metadata displayed on the OPAC interface is considered. However, there are two essential issues. Firstly, users want their metadata to remain intact, without being edited or removed by librarians or other users (a quality issue). This is because such metadata is relevant for their own use. Secondly, such inclusion of all user-created metadata results in large volumes of metadata (a quantity issue). Users also want to access relevant results without being affected by the large volumes of metadata.

Standards-based metadata systems do not permit user-led metadata creation and filtering. Thus users have no opportunity to manipulate their metadata view. This is due to the principles that underpinned standards-based metadata creation (2.6). In this situation, displaying only a few metadata values is not overly cumbersome. However, when metadata is enriched, displaying all metadata that surrounds an information object in a single global metadata view could confuse the user.

One interviewee noted that “if you have got a big load of metadata in the catalogue and also incorporate tags, including those idiosyncratic ones, that catalogue clutters things” (Study2-Interviewee6). Another interviewee concurred, stating that “the more you put, the more you also confuse” (Study2-Interviewee5). The interviewees suggested the resolution of this problem via the provision of multiple views, by dissecting metadata into various layers. The

consideration of multiple metadata views, the interviewees observed, should take into consideration the context of usage of each user.

“The challenge here is to present the results in a meaningful way. If I put my searches on [a particular search term], I may potentially get hundreds of results, and if we start to search newspaper articles as well as all the metadata, you might just get swamped. How do we then guide the customer and present the results of these many things in a meaningful way, without confusing them with the plethora of search results” (Study2-Interviewee5).

As mentioned earlier, contemporary OPAC interfaces provide a single global display to all users. One interviewee (Study2-Interviewee7) brought up the existence of tension between [metadata] simplicity of creation and the use and the display needs of various users. The interviewee maintained that some users require greater detail of metadata whilst others may prefer minimal description. Thus the interviewee criticised contemporary systems for presenting one global interface to all, without giving the user any control over the display. But, according to the interviewees, different users’ exhibit different display requirements, hence the interface should provide multiple views.

The Open Coding of interview data resulted in Open Codes such as ‘**filtering by users**’, ‘**incorporating socially-constructed metadata into the OPAC**’ and ‘**providing various views**’. The Focused Code that emerged from the analysis of the Open Codes is ‘**Providing Multiple Views**’ (see Figure 12.1).

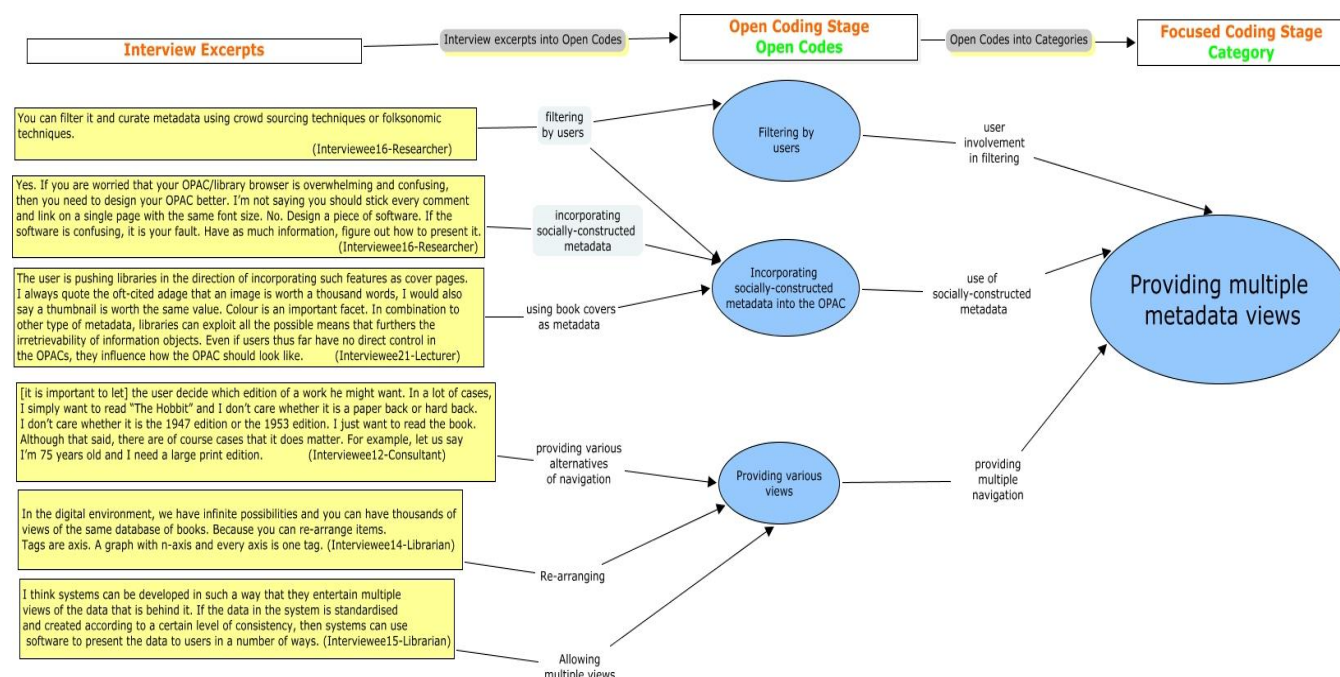


Figure 12.1: Providing Multiple Views

Current metadata principles are designed for a single metadata display. The interfaces are predetermined. However, in a mixed metadata approach whereby users are considered as pro-active metadata creators (8.4.2), and where there is metadata diversity (8.4.3), metadata

interfaces can be designed to provide multiple views (displays). The displays can be re-configured based on users' preferences. The filtering can therefore be tailored by the users themselves, providing post-hoc (on-demand) filtering.

12.3. Interface Simplicity

The Online Public Access Catalogue (OPAC) is the library's electronic front face that serves as one of the gateways for users to find and discover information objects of various genres. The OPAC first came in to use in libraries in the late 1960s. One interviewee recalled that the OPAC in its early days was mainly character-based on dumb-terminals; however, through the years the interface has been updated now incorporating some web-based features (Study2-Interviewee11). However, many interviewees hold the opinion that the changes have been minimal. The interviewees acknowledged the shift towards incorporating a Google-like search interface as well as the inclusion of book cover images and tables of contents. However, they also noted that, even these changes are not very welcome by all users or librarians. They indicated that, in the university context, undergraduate students seem to be relatively at ease with the simple single box search interfaces while post-graduates, researchers and librarians preferred advanced multiple box search interface that used to be the characteristic feature of the OPAC. For example, librarians, as one interviewee noted, are said to prefer the advanced multiple-box search options, due to the way the latter describes books.

“Because that is how they were taught, that is how they catalogued their items and it is not surprising that it is catalogued that way because that is exactly how the card catalogue evolved. Drawers and drawers of cards ordered by title, and then another cabinet of drawers and drawers of cards ordered by author, another block of drawers of cards arranged by subject. That is the way the items had been catalogued, that is the way the librarians wanted it to be and therefore the search interface reflected it” (Study2-Interviewee11).

One interviewee exclaimed “I don't particularly like the Google approach implemented in the library catalogue, but I do like it when I make a search on Google” (Study2-Interviewee6). The interviewee seemed to prefer the use of advanced search and browsing features implemented in contemporary library OPACs. It is however important to note here that librarians have the detailed knowledge of the intricacies of advanced searching, whereas users may not. Advanced search options require users to follow strict commands and search procedures which, as the interviewees believed, needed information literacy skills in using keywords and finding one's way through the various databases made available at libraries. In contrast, the single search interface allows users to simply type in keywords and results, if any, are presented on one screen.

The simple search box is being implemented as the default whilst advanced search is an optional feature in most present day OPACs. It is mentioned that interface changes were not always welcome by all users for two main reasons: they are used to the old interface and there is always a learning curve associated with the introduction of new ones.

The interviewees alluded to the fact that users prefer to use commercial search engines, rather than library catalogues, as their first point of search. As the interviewees noted, the importance of OPAC, as the main gateway to library resources, has declined. One of the reasons is that the interface lags behind the growing user requirements and expectations. In other words, as most interviewees highlighted, the library's OPAC is in competition with other web-based information providers such as search engines. Whilst, users may still use external search engines to retrieve a library's collections, it is no guarantee that all the library's collections are findable and discoverable via search engines. Interviewees indicated the importance of adding flexible searching options, such as 'Did You Mean' feature, in the OPAC.

In relation to interface simplicity, the interviewees identified the following issues that need to be addressed. These include, firstly, the need to determine users' requirements. With regard to this, one interviewee (Study2-Interviewee21) noted "The user is pushing libraries in the direction of incorporating [new metadata] features. Even if users thus far have no direct control in the OPACs, they influence how the OPAC should look like". Secondly, the interviewees indicated the importance specialised expertise in interface design. It is indicated that currently libraries do not have the sets of skills required to design, develop and maintain a library metadata system interface that meets the needs of their users. For example one interviewee noted that "We are not Apple Inc., we don't have an expensive design team, it looks pretty good. So if your system is confusing, your software is broken and you need to fix it." (Study2-Interviewee16). Thirdly, cost implications of interface design were indicated. As one interviewee (Study2-Interviewee12) indicated "it is about balancing the requirements of the vendors and the customers because change will require development effort and it costs money somewhere. It either costs money for every library to develop those features locally, or it costs money for the vendors to do it". Figure 12.2 shows the result of the Focused Coding of interview data which resulted in the Category of Interface Simplicity.

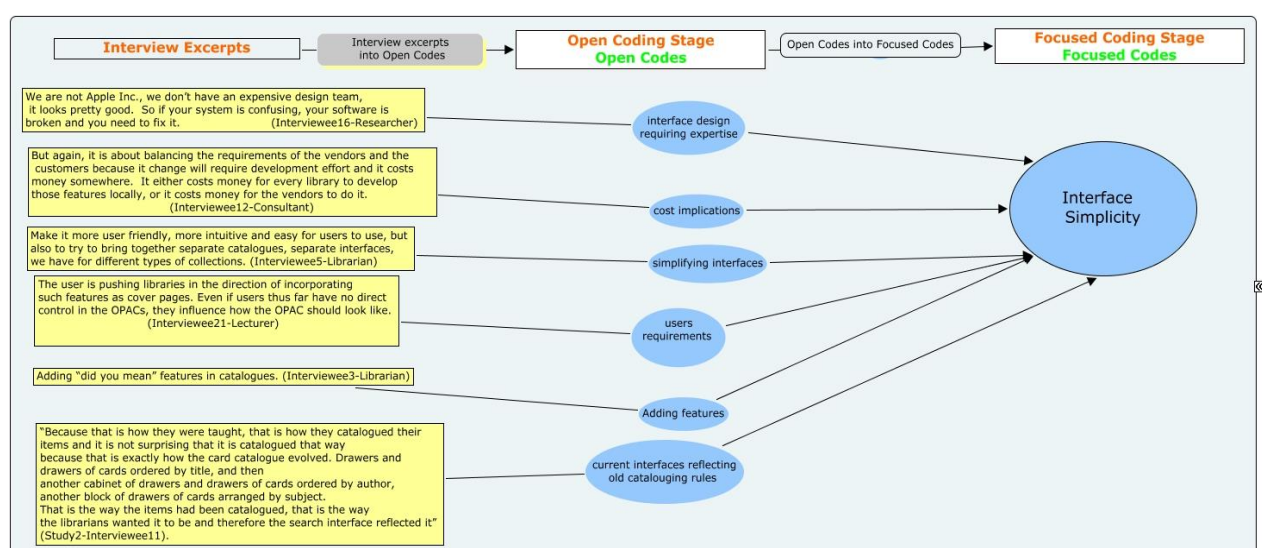


Figure 12.2: Interface Simplicity

The interfaces of current library OPACs, as the interviewees indicated, essentially reflected this quest for metadata simplicity. It is argued that the notion of metadata simplicity has been miss-interpreted by libraries, as they focus on simplifying and reducing the number of metadata fields displayed on the OPAC, rather than focusing on interface simplicity. However, in light of the future of metadata that can be enriched through proactive engagement of users as co-creators, there is an apparent need for a re-conceptualisation of the principles of contemporary metadata approaches. Thus, in this thesis, interface simplicity rather than metadata simplicity, is suggested as this refers to usability, which is chiefly a technical issue.

12.4. Personalisation

The interviewees noted that personalisation is a contentious issue, especially in view of its implications for the privacy of users. Personalisation and privacy (see Figure 12.3) are considered as highly interdependent, as the former can only be effectively implemented by using personal user data.

“The issue is you need to know an awful lot about the user and that perhaps raises privacy concerns which aren’t instrumental but you would need to engage in an open and frank conversation with the users to help them understand how the data will be used, where it will go, who can find out what about them, and also the value it gives them in return. Perhaps you are looking at an institution entering into a contract with the user and say you give me the following data about yourself, it will be used in these ways, and you will get the following benefits in return, you either opt in or you don’t” (Study2-Interviewee12).

Collecting more information about users also implies that libraries are taking legal and ethical responsibilities to maintain the privacy of the user. Thus, there should be a balance between the two. Interviewees suggested that user profiles can be used without disclosing the particular details of an individual user, it can be used anonymously. As one interviewee indicated, generic information with regard to the interests of specific groups of users, such as students taking a particular course, has little or no implications on privacy; thus, such types of personalisation can easily be affected by libraries. The interviewee indicated this by saying that “[If you] know that first year PhD students in the physics department borrowed the following books, you are beginning to get close to personally identifiable information. Once you get that close, you need to start these conversations with the user. And you do it very frankly. Tell them what the benefits are. Tell them what they will get with or without that information” (Study2-Interviewee12). As the interviewees indicated user’s search and navigation histories can provide an important source of information to cater for recommendations, according to their requirements. However, the privacy and legal concerns and stakes of such types of information are higher.

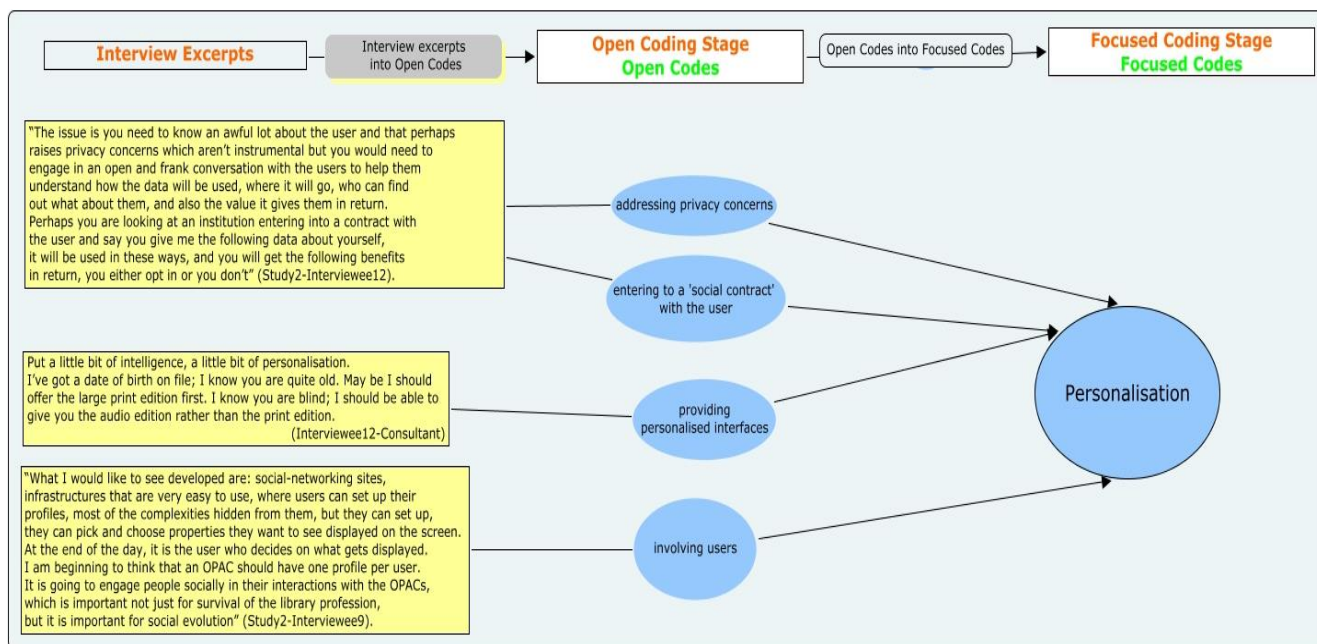


Figure 12.3: Personalisation

Effective and careful usage of user profiles is considered an important component of personalisation for metadata presentation.

“What I would like to see developed are: social-networking sites, infrastructures that are very easy to use, where users can set up their profiles, most of the complexities hidden from them, but they can set up, they can pick and choose properties they want to see displayed on the screen. At the end of the day, it is the user who decides on what gets displayed. I am beginning to think that an OPAC should have one profile per user. It is going to engage people socially in their interactions with the OPACs, which is important not just for survival of the library profession, but it is important for social evolution” (Study2-Interviewee9).

Metadata personalisation can help implement multiple views, catering for differences in the individual users' needs in a better way than the traditional single global view. For example, users would be able to navigate through their own preferred catalogue of books, articles, tags, comments, reviews, and favourites, which they had saved previously. In addition, metadata associated to user's search histories and browsing and navigation histories can also be saved and be used for personalisation. Some interviewees also mentioned the potential benefit of circulation data for tailoring, i.e. personalising, displays. It is important to note that less user personal data might imply less personalisation, which would also affect the quality of the presentation of metadata in multiple views.

The interviewees were cognisant of privacy issues when using user profiles and thus they underscored the importance of negotiation and consensus with users in using such data. Users will necessarily want to know how the provision of their personal profiles would benefit them, in relation to their experiences in finding and discovering information resources. User profiles might as well be used to power recommendation services.

12.5. User Recommendations

Users who come across an information object might suggest or recommend it to their peers and colleagues. Recommendations can be communicated through various means, including email, messaging, social networking and other avenues of communication. The recommender is expected to have some prior knowledge of the person to whom the recommendations are being made.

“As an undergraduate in 1968, I remember, a particularly interesting course we were doing in which the text book was not very good. However, purely serendipitously, a friend of mine was browsing in the university bookshop and he came across this newly published book and he said, ‘well, this might be written for this course’, and everybody bought a copy. That book absolutely might have been written to meet the demands of this course that we were struggling with, [better than the recommended text as it was] a bit out of date and a bit difficult. So we all bought copies of this new book” (Study2-Interviewee1).

For librarians, to implement recommendation services, they need to have some basic or detailed profile of the user. The more information they have, the more specific and relevant the recommendations should be. As one interviewee (Study2-Interviewee17) suggested, users’ profiles are important when incorporating user recommendation features in library metadata interfaces. The interviewee said “I think if we can have that kind of contextual display where the system detects the location of the searcher. In addition, for example using a login system, the system can be identified from which department he/she is using, suggestions they can get, so that the recommendations they get are catered to their field of study”. Thus, libraries can develop networks of interest around particular user groups so that recommendations can happen among the users themselves. User recommendations features imply that users can now collaboratively filter information objects.

12.6. System Recommendations

System recommendations refer to the use of automatic means to filter what are considered relevant information objects to users, based on their profiles, browsing history and search experiences. Most interviewees mentioned Amazon.com as a good example of filtering through system recommendations. They alluded to the fact that automatic recommendations may not always be accurate but play a role in the filtering process. System recommendations may facilitate serendipitous discoveries (see 12.8). Figure 12.4 shows a simple relationship between user profiles (educational level, age, language and personal interests), searching, discovery and recommendations.

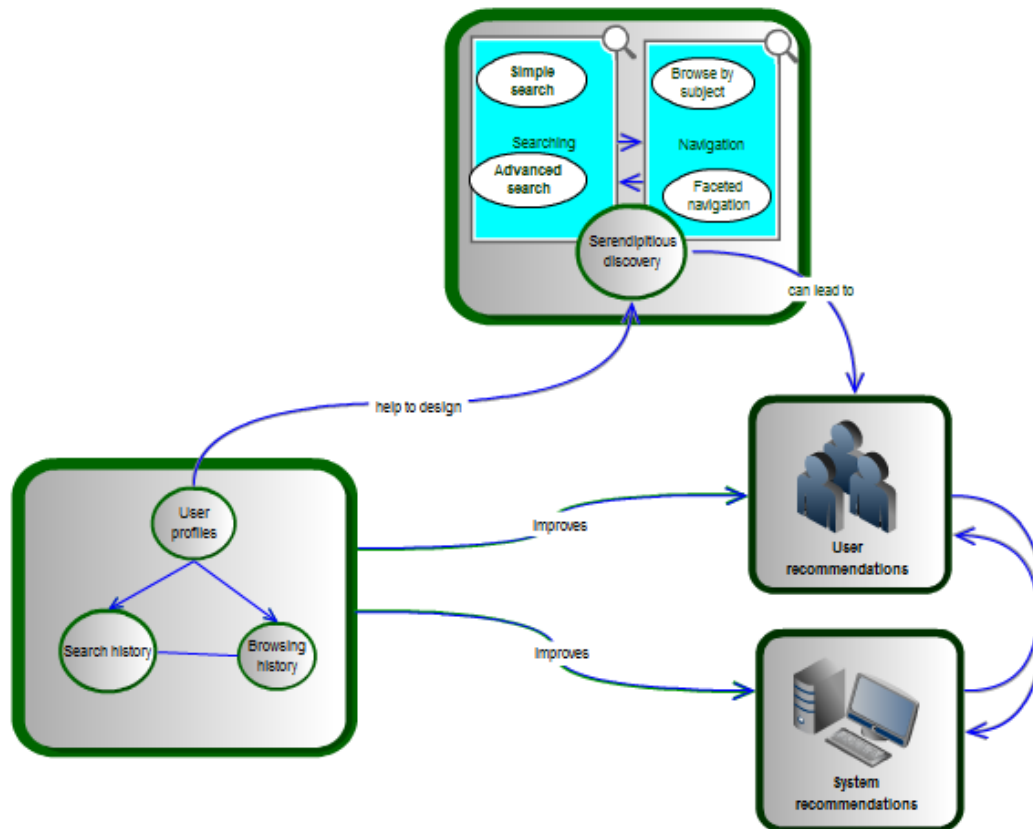


Figure 12.4: Using User Profiles for Recommendations

Recommendations, both system and user, are about catering to the needs of users. Recommendations are dependent on personalisation (personal profiles/interests), thus indicating the importance of addressing privacy concerns of users (see Figure 12.5).

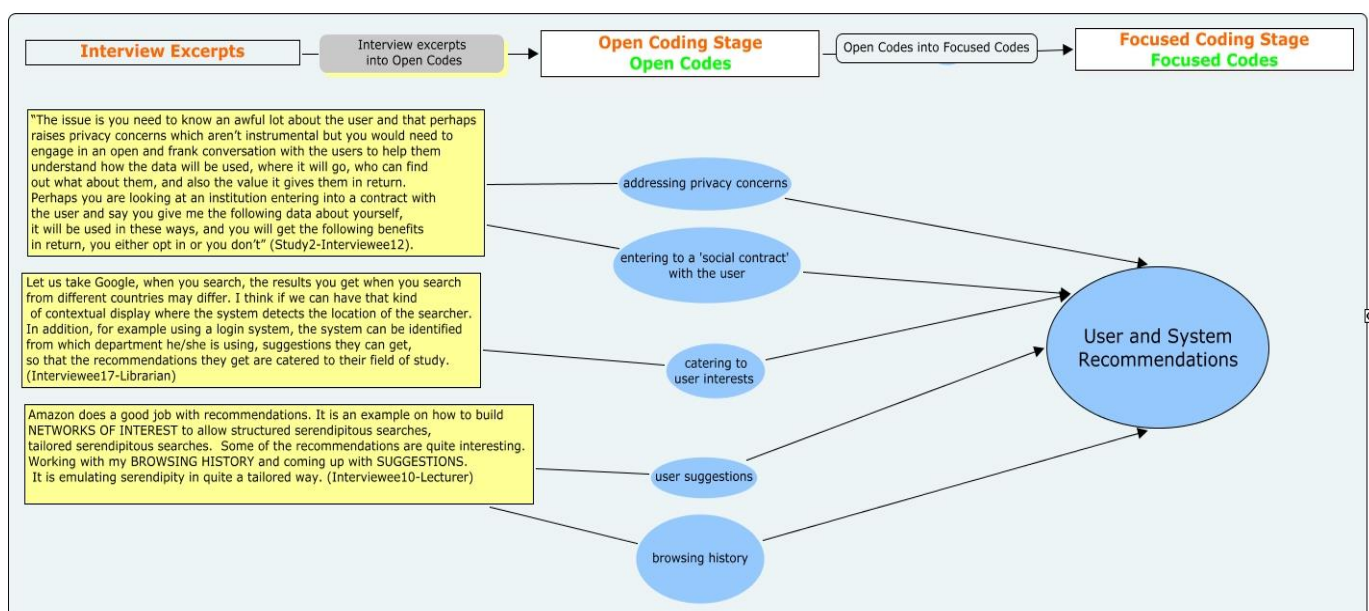


Figure 12.5: User and System Recommendations

12.7. Integrated Access to Electronic Resources

There is consensus among interviewees that contemporary library OPACs are unable to provide integrated access to the various genres of collections in libraries. Users are often forced to search separate databases. The problem is partly attributed to the fact that most of a library's databases are proprietary and are designed differently. One interviewee noted that "The fact is that every database we have is different. There is no one single platform. Well, there are some databases who try to gather a couple of sources together, but IEEE looks nothing like Science Direct. It just shouldn't be so, so I am hoping that will change" (Study2-Interviewee2). Libraries have no access to source codes or even the expertise to manipulate the codes, in order to be able to integrate the various interfaces. As a result users often seek librarian assistance to access the various databases and search through specialised interfaces. Hence, one interviewee (Study2-Interviewee1) remarked that "the OPAC could be much better", if it seamlessly integrated all the resources the library makes available.

"I would desperately like to hope we have much more kind of joining up, a bit like what you were saying. The whole thing of for example Shibboleth login, the technology there, you can see why it is there protecting copyright, it is just so massively horribly complicated for the user and the likes of myself to explain to the users. I desperately hope that in the near future, that will become more straightforward. I would really like to hope, it would become much easier for students to say I am interested in this subject; where do I find information? At the moment, I have got to search this database, and this database and that database, Google Scholar, the Library OPAC, or whatever" (Study2-Interviewee2).

Other interviewees also concurred, stating that integrated search and navigation of disparate databases is highly sought after by their library users, which nonetheless remain unfulfilled by the OPAC at present. The interviewees expressed their wish to see the OPAC developed into such an integrated discovery interface, whereby the resources which are currently locked in proprietary systems are integrated into searching services including books, journal articles, institutional repositories, and abstracts (Study2-Interviewee11). The Open and Focused Coding stage of data analysis resulted in the Focused Code of 'Integrated Access' (see Figure 12.6).

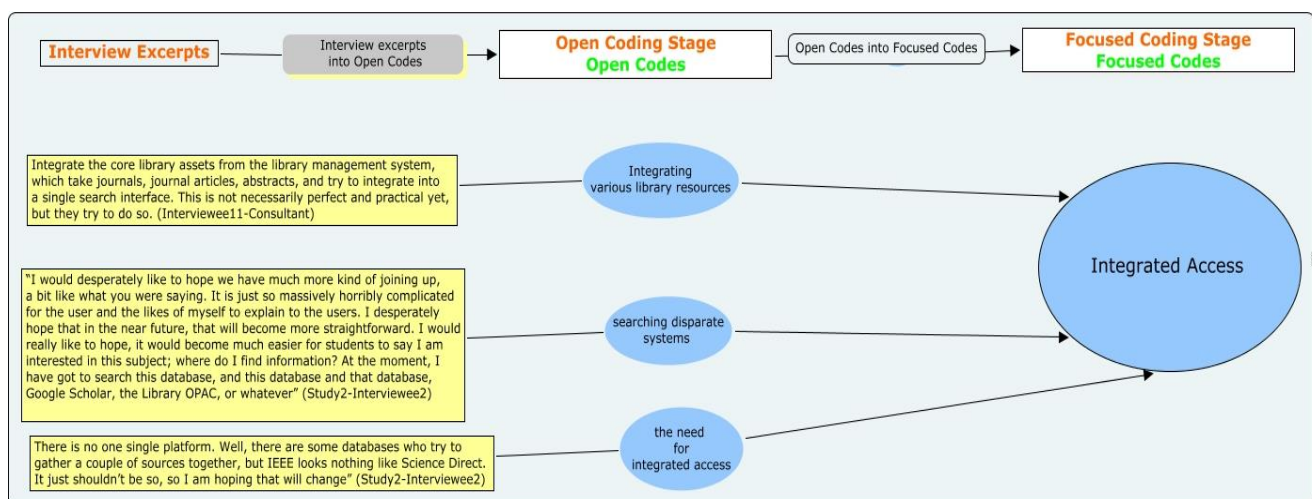


Figure 12.6: Integrated Access

12.8. Facilitating Serendipitous Discovery

Interviewees contrasted browsing physical library shelves with its online (digital) counterpart, especially as to how the latter is implemented in contemporary library catalogues. As one interviewee indicated “digital libraries are losing some of the serendipity. I am not sure how much serendipity is there these days” (Study2-Interviewee6). The interviewee, who is a senior librarian, stated that, compared to his student days, when there were no electronic information services, there was much more serendipity than at present. Another interviewee said that “there is a bit of me that hates the virtual because you lose serendipity” (Study2-Interviewee2). It is indicated that serendipity is an important component which “gets people to think outside their purview” (Study2-Interviewee8), hence suggesting that it has wider implications than mere reading. In other words, serendipity may help chart new research areas, the exploration of inter-disciplinary issues. As observed by one interviewee, serendipity is helpful for discoveries outside the remit of one’s area of interest. The interviewee stated that “Many problems require the expertise of several disciplines, and digital libraries are one of them. We need to acknowledge there are these natural links between disciplines” (Study2-Interviewee8).

Serendipity is therefore an important issue that should be considered in connection with metadata approaches. Digital libraries, by their very nature, tend to encourage known-item searches, whilst physical libraries require effort to go around shelves, forcing users to take time to find, identify and retrieve information objects, thereby, at times, facilitating fortuitous discoveries. As the interviewees indicated physical libraries have a built-in design for serendipity. Hence, the limitations that it imposes on users, to browse shelves may at times result in unexpected but interesting reads, leading to new research and study focus. The interviewees identified the challenges of emulating serendipity in digital libraries, whilst at the same time indicating the possibilities of designing discovery systems with such possibilities of serendipitous access. Serendipity is partly therefore an interface (application) issue.

The interviewees indicated that serendipitous discovery of information objects widens the possibilities of finding diverse and often unsought solutions to problems. As one interviewee noted, undergraduates that they have observed have a preponderance to focus on reading lists given for their courses by their lecturers and much reliance on predefined items. The interviewee contended that such behaviours have increasingly made students care less and less for serendipity. Most interviewees maintained that the lack of serendipity in digital libraries is partly attributed to the preoccupation on search, rather than navigation and discovery systems. Pointing to the known-item search focus of current library catalogue designs, interviewees suggested the importance of designing serendipity in library systems. One interviewee noted that “serendipity is important. It will always be important. How you replicate that, the serendipity you get from a physical book shelf, is an open question” (Study2-Interviewee12).

Many interviewees, especially, librarians stated their experience was that some users come to the library with a general research area, thus they may not know what they are specifically looking for. Thus, inducing serendipitous discovery into digital libraries would play a paramount role. The interviewee noted that “See also” references, associative trails and links can enhance such serendipitous discovery to information objects (see Figure 12.7).

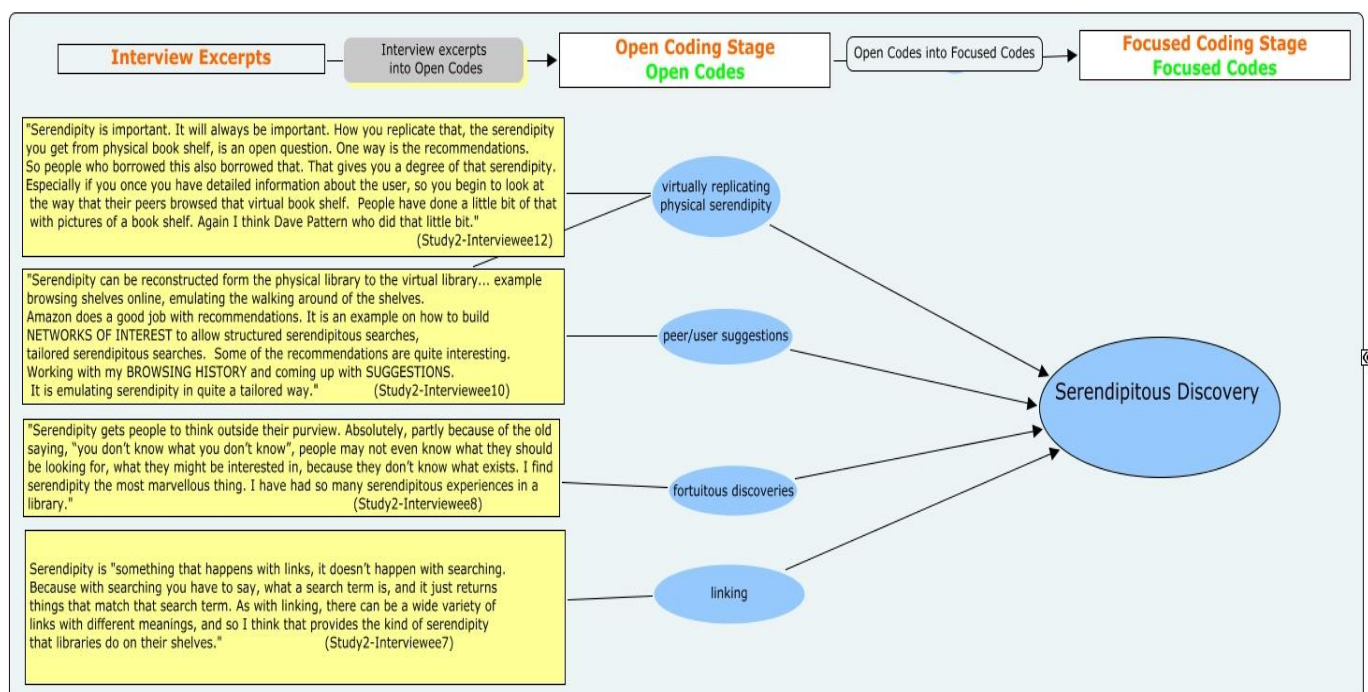


Figure 12.7: Serendipitous Discovery

Open access to physical library shelves tend to make serendipitous discovery of a book or a journal possible. Those libraries that enforce closed access to the stacks of books offer very little in terms of such accidental discoveries of books on shelves. OPACs, which are designed to answer predetermined item searches (e.g., using author's name or title), are analogous to closed access practises to books on physical library shelves, as both rely on matching users' queries with records available in the collection when responding to a user's quest for an information source. Put simply, the element of pleasant surprising discovery which could be

realised by physical library shelves could altogether vanish if digital libraries were to solely focus on matching user queries. This then begs the question, ‘how can such serendipity be incorporated in the design of digital libraries?’ Since access to information objects in digital libraries depends upon the use of metadata, part of the solution lies with the choice of metadata approaches. A mixed metadata approach provides enriched metadata from users and such metadata can be filtered by and to users using emerging metadata platforms such as tag clouds and recommendations, improving therefore the likelihood of serendipitous discoveries.

12.9. Conclusions

From the data analysis, metadata filtering emerged as an important core category wherein libraries and system developers should consider in approach to discovery interfaces. The emergence of the Core Category of metadata filtering entails the importance of providing contextualised metadata views based on individual interests.

In a mixed metadata paradigm, metadata better conforms to the multitude of perspectives and interpretations of various groups of potential users. The size and diversity of metadata increases as users continue to create more metadata. In order for users to effectively utilise the metadata and thus enhance the findability and discoverability of information objects, efficient metadata filtering is essential.

Current metadata interfaces are designed a priori and they provide a single interface for all users. Furthermore, as a consequence of a purely a priori and expert-led design and metadata creation, fewer metadata fields are encouraged. However, reducing the number of metadata elements removes from information objects some of their descriptive power. Unlike the single global interfaces available in current library management systems, a reconfigurable metadata interface supported by users’ profiles (personalisation) is suggested (Figure 12.8).

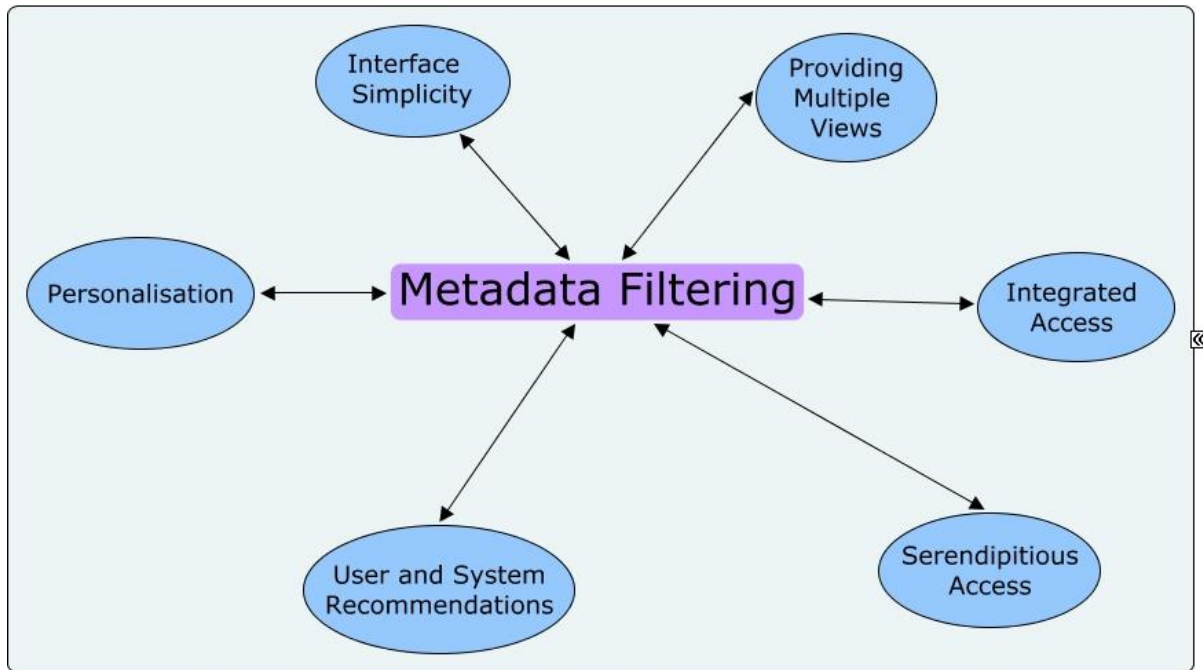


Figure 12.8: The Core Category of Metadata Filtering

Put another way, metadata enriched through the combined use of standards-based and socially-constructed metadata approaches should be filtered and presented to the user whilst supporting their individual information needs, rather than an a priori generalisation. Such interfaces should also support serendipitous discoveries in addition to known-item searches. The principle of metadata filtering espouses the importance of delineating metadata content and presentation. It introduces post-hoc metadata presentation based on users' needs. Informed by a mixed metadata paradigm, the principle states that metadata filtering should be continually enhanced through user-led, re-configurable and customisable user interfaces. In order to effectively incorporate users' preferences, it was suggested that the principle of metadata filtering is informed by users' needs. In relative terms, the interviewees dwelled in more detail when they discussed metadata enriching. But they lack detail in their discussion on how filtering can be improved. Thus, a grounded theory analysis of users (Chapter Thirteen) is required to saturate the category of metadata filtering.

Chapter Thirteen: Study Three - Analysis of the Perspectives of Library Users on Library Resource Usage and the Role of Socially-Constructed Metadata Approaches

13.1. Chapter Overview

It was indicated that metadata that is enriched (Chapter Nine) is also required to be linked (Chapter Ten) and openly accessible (Chapter Eleven). The emergence of these Core Categories was mainly derived from analysing the perspectives of Library and Information Science professionals (see chapters Seven and Eight). In addition, metadata that is enriched, linked and openly accessible, as discussed in Chapter Twelve, is required to be filtered and presented to users, thus the emergence of the Core Category of metadata filtering (see Chapter Twelve). However, it was also indicated that this needed to incorporate the perspectives of library users. Thus, Study Three was designed and executed (Chapter Thirteen), and used to saturate the Core Category of metadata filtering.

During this stage, in conformity with the Constructivist Grounded Theory Methodology, Theoretical Saturation was employed for refining, extending and saturating the categories that had emerged from Study Two. Furthermore, novel concepts and categories were also allowed to emerge from Study Three, with the intention of enabling the study to fully incorporate the perspectives and views of library users. The interview for library users paid particular attention to the library catalogue, i.e. the library website/the interface/OPAC. In light of the core categories that had been identified in Study Two (metadata enriching, linking, openness and filtering), the analysis of Study Two from the perspective of Library Resources Users, hereafter referred as to Users, focused on the metadata filtering. Thus, the analysis of the data on User interviews is presented immediately after Chapter-Twelve: Metadata filtering. Employing an iterative data collection technique, a total of 25 interviews were conducted with library users including students (both undergraduates and postgraduates) and university lecturers (see Table 13.1). As the perspectives of LIS professionals have been covered in studies One and Two, this study explores only the views of users who were not affiliated with the field of Library and Information Science (LIS).

Study No.	Educational Level	Sub-total	Total
Study Three	Students – Under-graduate (BSc/BA)	4	25
	Students – Post-graduate (MA, MSc)	4	
	Students – Doctoral (PhD)	6	
	Lecturers	11	

Table 13.1: Interviewees for Study Three

Based on the findings from the Constructivist Grounded Theory analysis of Study Two, semi-structured questions were prepared to explore and investigate the perspectives of Users (Appendix-1.3; Appendix-2.3). Interview questions mainly focused on the interfaces to library websites and catalogues, as this was deemed most crucial to this group of interviewees. Furthermore, concepts that fell under the core category of metadata filtering were employed to extend the scope of the subsequent discussions and, hence, the data analysis. However, in order not to constrain the emergence of novel concepts, discussions were intentionally left open-ended. The objective of the interview was to explore users' experiences in finding and discovering information objects for teaching, research or in-depth studies. For example, questions were identified how users go about finding information: Where do they start their search for information? Do they go to the library physically or do they use it online? Questions also addressed users' experiences in using the library catalogue (OPAC) as well as their views regarding the usefulness of the metadata (information) displayed in search results. Users were also asked to describe their perspectives on the potential role of social media and Web 2.0 approaches in libraries. In particular, they were asked specific questions on their experiences in the use of online book reviews, ratings, tagging and recommendations as examples of instantiations of these concepts.

Twenty-two interviews were conducted face-to-face while three others, due to distance constraints, were carried out via Skype. In addition, all but one interview (at the request of the interviewee) were audio-recorded and transcribed. Two stages of the Constructivist Grounded Theory Methodology, namely Open Coding and Focused Coding were adopted whilst NVivo 9 was used to support qualitative data analysis.

13.2. Research Narrative of Interviewees' on the Current Status of Library Systems

Before discussing the Categories identified to saturate the Core Category of metadata filtering, a narrative of interview discussions is given in the following section. The interviewees indicated the changing information services landscape by discussing the changing role of the library from physical reading place to a virtual information space (13.2.1), the importance of re-orienting the Online Public Access Catalogue to meet users' information needs (13.2.2) and the role of the web as the first stop for information (13.2.3).

13.2.1. The Library as a Physical and Virtual Space

As the Focused Coding of interview data showed, most interviewees agreed upon the continued relevance of libraries as well as the collections they hold and the services they rendered, whilst accentuating the need for adapting to changing users' needs, especially with the increase in demand for electronic access to books and other information sources. Libraries as physical spaces, most interviewees agreed, will continue to be relevant, even though, due to time pressures and other constraints, they are less likely to visit libraries in person. As a result, lecturers, researchers and PhD students seemed to mostly rely on electronic access to library resources rather than using libraries as a physical collection or reading space. On the other hand taught students use the library as a study place. However, they revealed that they

were not entirely dependent on any particular library's website and its catalogue, mainly due to the continued availability of digital information, including library databases, inter-library loans, use of commercial search engines and the possibility of purchasing books online.

In regard to this, one interviewee (Study3-Interviewee24) remarked that "I would never want to miss the opportunity of visiting libraries and enjoying the smell of books. But we have the realities of restrictions of time, and meeting all our deadlines". As a result, the interviewee foresaw reductions in the frequency of library visits. They added that "I very rarely physically go and use the library unless I exactly know what I am looking for. I'll exhaust every possibility before I physically go to the library." The Interviewee, however, further asserted that their information needs are better met now than previously, mainly as a result of the emergence of electronic access to library databases, e-books and various alternatives to full-text information.

One interviewee (Study3-Interviewee14) indicated that they have not been to the library for several years, partly because the courses they taught were computer-based and the requisite resources were accessible electronically. But this did not necessarily mean that other lecturers, even in the same subject domain, did not use the library. For instance, one interviewee (Study3-Interviewee11) asserted that they often visit the library to borrow books. However, the latter interviewee also noted that they had other access avenues to books: review copies from publishers, second-hand purchases from Amazon, as well as others.

Yet another interviewee noted the unique nature of libraries as a physical space, mainly due to their solemnity, stating:

"I like libraries as a brick and mortar place, to just go sit down and work. I don't necessarily go to the library to use the books. But I could go to a library to sit down and use the Internet connection and benefit from the quietness associated with that" (Study3-Interviewee15).

The same interviewee, however, stressed the importance of re-orienting libraries in light of changing needs and evolving trends. The interviewee emphasised the need for moving into the provision of electronic access to information as well as for adopting flexible use policies, and for attracting users. The interviewee further asserted that "I think libraries should embrace mechanisms that make it easier to see the value in the knowledge-base they are holding onto. They have shelves and shelves of books, which may or might not be useful, and these books are not accessible".

One interviewee (Study3-Interviewee9) mentioned their practice of initially trying to gather as much information about a particular resource from a search engine and then physically walking over to the library to borrow a book or a journal. Depending on the visibility of the library collections on such search engines, library resources, such as electronic journals, are displayed. One interviewee (Study3-Interviewee13) indicated that they found the Internet a more expedient place to start their search, asserting that "it is so much more convenient, rather than going to the library". This interviewee further attested that they utilise library collections; nonetheless they preferred electronic versions of these collections.

Thus, students tended to prefer libraries as a place of study where they benefited from its quietness, the opportunity for browsing through the collections, and getting access to computing resources and Internet connections. However, lecturers and researchers, who tend to have access to the latter resources at their own offices or laboratories, do not have particular interest in sitting in libraries. However, all groups of users tended to appreciate the continued importance of libraries, both as a physical place and a virtual space.

Interview data also revealed the importance of availing information in both physical and digital formats, depending on users' needs. Users emphasised the importance of giving access to digital information in libraries, including e-books, in a format that is convenient and easy to use. One interviewee noted: "I tend to actually use the library [resources available] online, much more than going into the library and finding a book. I like to find it online so that I know they've got it. Then I like to know where it is placed in the library, rather than having to go to the library and finding out what section it is in; it is a confusing system. I can't do that" (Study3-Interviewee1).

Furthermore, some interviewees noted that as their courses tended to heavily rely on online resources, getting access to computing resources, including desktops, laptops, tablets and mobile devices, at all places was crucial, demonstrating a requirement for having access to information 'on the go'. An interviewee (Study3-Interviewee1) provided details of their information discovery strategy, stating that they start a search for factual information on a search engine, and then use databases when searching for academic information. Whilst they access the library website once in a while, they found it easier to use general-purpose search engines and external databases. Due to the nature of their course (Computer Games), this interviewee indicated that they want to use the library only if it is easier to use and access than other alternatives.

13.2.2. The Current Status of Library OPAC

When it comes to access interfaces, i.e. the library's website and/or its catalogue, there seemed to be widespread agreement that these were found to be too difficult to use and/or that one had to go to induction training, if one wanted to self-access the library's resources. Whilst it was established that users in general tended to start their search for information from search engines, such as Google or Google Scholar, these would ultimately, depending on proxy settings, lead them to full-text links to resources that are subscribed to by the local library, if any.

It is indicated that the provision of continuous, user-friendly, persistent and pervasive access to a library's website is crucial, if the latter is to accomplish one of its mandates. For instance, it is important to make the library website easily accessible through a university's website as well as through its Learning Management System, thereby enabling a seamless navigation setup across the various resources (access points). For example, whilst some interviewees started their search directly from the library website, others are led to it from general-purpose search engines, such as Google or Google Scholar. When searching on these general-purpose search engines, based on automatic and proxy settings, the resultant search retrievals indicate whether a given resource is available at the local library. Users seemed to consider this a very

fast and easy way of accessing full-text articles. When it came to books, users seemed to prefer searching the OPAC. For journals, they tended to prefer search engines over the library catalogue, since the latter does not provide them with an easy means of retrieving full-text articles.

With regard to the library interface, several users found it cumbersome; making searches a chore, especially when compared with doing so through either Google or Google Scholar. In general, the library OPAC was considered as both unappealing and less user-friendly than general-purpose search engines, with which users are familiar. Some of the recurring concerns of users included a lack of speed, number of clicks required, steep learning curves, and an absence of immediate access.

As the interview data suggested, it is crucial that libraries build novel services around their documentary collections and other information resources. As lecturers, researchers and students need to gain access to these resources at their own convenience, using their personal devices, including PCs, laptops, tablets, e-book readers and mobile phones, it is imperative that libraries continually re-orient their services based on current and emerging trends. Users might also want to access resources on the go, making the provision of digital versions of these resources essential.

It was also pointed out that e-book databases available in the library were not convenient for use on user's personal devices, as each had its own search interface, which made it rather difficult to use. While users mainly wanted to read these e-books, they felt impatient at the intricate set of requirements for using them. Thus, they often opt for other providers, which enabled users to use fewer keywords and find results.

As users seemed prone to bypassing the OPAC, and as a consequence bypassing also library e-resources, mainly due to the issue of usability, libraries should not only re-conceptualise the essence of an OPAC, but should also ensure that their resources are discoverable through commercial and other general-purpose search engines. Hence, the emphasis should be on access, rather than the library brand. For rendering an optimal service to users, it should not matter whether a user accesses the library e-book or journals collection through Google or other search engines. Libraries should focus on making their resources discoverable and findable. This clearly demonstrates the need for linking (Chapter Ten) and openness (Chapter Eleven) as previously suggested.

On the other hand, it is important to note that books, both print and electronic, are deemed relevant. One interviewee noted that there are people who are genuinely interested in reading books. However, if such books were cumbersome to find and are not easily accessible, users may opt to find alternative ways of accessing them. Some users find it more convenient to buy books online, whenever they can afford to do this. The value of libraries is therefore becoming closely tied to their competence in enabling users to find books and other resources available in the library and to access them immediately.

During discussions on the future of libraries, especially in relation to opening library websites for enabling users to contribute metadata through reviews, tags ratings and recommendations,

one interviewee (Study3-Interviewee15) pointed out the importance of having trust in users. The interviewee believed that, at present, libraries are overly concerned about malicious entries and corruption, if they opened their library systems for users to co-create metadata. Such fear and lack of trust, the interviewee noted, is likely to drag libraries out of the information systems picture, as users gravitate to more open ones. The interviewees indicated that libraries should aspire and adapt to changing circumstances. This includes permitting the participation of users in co-creating metadata.

13.2.3. The Web as First Stop for Information

The interview data suggested that all library users were aware of the existence of a burgeoning variety of digital information across the Web. The interviewees also noticed an increasing predisposition by libraries to collect a diversity of sources, such as electronic databases, e-journals and e-books. Due to changes in technology, as the interviewees acknowledged, their reliance on the web as the first stop for information has risen progressively over the years. Increased access to computers, widespread availability of connections to the Internet as well as convenience and immediacy of finding information on the web, the interviewees maintained, played a pivotal role in their increased dependence on the web. Interviewees also acknowledged the importance of the library in providing access to information resources.

The interviewees were emphatic in stating that they, more often than not, tended to rely on the generic web search engines when starting their search for information. However, a closer examination of the interview data revealed that they in fact use library resources heavily, which are made available through automatic proxies (e.g. Shibboleth login). As it happens, when users are connected within a university network, their searches on general-purpose search engines, such as Google or Google Scholar, retrieve results that are accessible through a university's subscriptions. Thus, whilst undergraduates seemed to make no distinctions as to where the retrieved results come from, researchers and lecturers are relatively more cognisant of the fact that most of their article searches are served up by the library, even if they were not using the library catalogue (OPAC) as their first point of entry.

While relating their experiences with the web, most interviewees expressed their preference for starting their search for information on the web. This is a trend that is likely to become more prevalent, according to one, remarking "I would now start with the web. If it were four or five years ago, I would have gone to library texts" (Study3-Interviewee12). In comparison to the library and subject-specific electronic databases, users seem to prefer using search engines as a first gateway to the web. One interviewee (Study3-Interviewee1) noted "Google is my first step" and especially noted the ease of use of this search engine: just typing keywords on the address bar, without even the need for keying in the web address of the search engine. This and similar changing requirements of users, when accessing diverse information sources, have significant consequences on the way library catalogue interfaces are designed as well as the way their services are offered. One impetus for change, for instance, is the fact that users are hard pressed for time and, hence, the need for sifting (filtering) through the relevancy of the retrieved information. Some of the factors

interviewees noted regarding their preference of the web as their first stop to information include: convenience of access (13.2.3.1), instantaneous access (13.2.3.2), up-to-date-ness (13.2.3.3), access to full-text (13.2.3.4) and about-ness metadata (13.2.3.5).

13.2.3.1. Convenience of Access

Users expressed their continued preference for convenience of access to information sources and services. They tend to start their search for information at the closest source from which they can easily access information, preferably from their armchairs, without the need for making any extra effort, when given the choice. All interviewees noted that they start their search from the web. One interviewee (Study3-Interviewee15) emphatically expressed his preference, stating: “I am a Googler”. Being a lecturer, the same interviewee further discussed employing multiple information sources. Interviewees seemed to demand still more, as exemplified by one interviewee’s remarks:

“What I would like is a library website [which], when I go to it, knows who I am, knows what I teach, knows in which department I am in, and already starts targeting the type of information resources [that] I want. What I do want is an interface that, when I come in, says the last things you looked at were ‘these’; other people looked at ‘these others’” (Study3-Interviewee14).

The same interviewee also stated their preference for library services that are targeted at their information needs. The interviewee stated that “If I go to the faculty librarian directly, she recognises me, others don’t. When I go to the library website, if it were configured for me, it would be more helpful for me. I want to see targeted resources, such as journals”, further indicating the importance of using profiles and personalisation (see 12.4) rather than each time asking to fill-forms or asking the same query over again. In their own words, the interviewee stated “I don’t want to go to the library website, login and fill-in a form; it shouldn’t be that way at all. The profile should go from a staff’s portal. As I click through, it knows who I am, it links to Moodle, it knows the courses I teach” (Study3-Interviewee14).

The above seems to succinctly summarise the concerns of most interviewees and also their needs. Another interviewee (Study3-Interviewee1) added “A perfect library for me would be one that automatically looks out for and retrieves a resource for me, but I think that is too much to ask”. Furthermore, in relation to convenience of access, most interviewees seemed to consider the current library systems rather obsolete, especially when compared to popular search engines. Users were also prone to criticising library catalogues as inconvenient. Notwithstanding the above, they equally recognised the importance of using library collections, as these lodge relevant resources that cannot be acquired by users individually.

13.2.3.2. Instantaneous Access

Speed of access to information is one important factor for users’ preferring the web as their first stop to search for information. General-purpose search engines, such as Google, seem to have excelled at providing instantaneous search results, at least much faster than those provided by libraries that, as some interviewees noted, took a longer route to the specific full-

text contents. Most interviewees associated this with a greater number of clicks and, hence, the browsing time it takes to access information. Furthermore, most interviewees remarked that they found the library website very complex, as it requires them to go through various paths and browsing hierarchies, with a few stating that they were unable to search for articles in journals listed in the current library catalogue. Almost all seemed to prefer a simple, Google-like search interface. Some users suggested using a maximum of three-clicks (Study3-Interviewee1; Study3-Interviewee2).

“For example, for me to access a book or an article, the library website takes 10-12 clicks at a minimum. But, I would say 2 to 3 clicks [would be ideal]. So that you type what you want, in the same way you do in Google Scholar. In Google Scholar it is a one-step operation. If it does not suit me, then I can try something else, of course. I don’t have to choose this, choose this, then choose that, yet make another choice, and at last find nothing at all” (Study3-Interviewee2).

From the above, it is evident that interviewees consider immediacy of access an important concern, suggesting that the number of clicks that it takes to access full-text of documents be reduced to a minimum. Moreover, users seemed to be even more frustrated when, after spending substantial time conducting a search, they were unable to retrieve full-text resources, as the library did not have a subscription, or when the items retrieved were not relevant. Users tended to prefer popular search engines, mainly because the latter provide results instantly; hence, the precision of the search mattered less as they were willing to try many combinations of search keywords, where the lack of precision seemed to be compensated by the speed of retrieval. As results are displayed during the search process, including when typographical errors are made, it gives users the leeway to try many search operations until they get an optimal retrieval.

In other words, the more quickly a system displays results, whilst providing flexibility for users to experiment with their search strategies, the better it resulted in user satisfaction. Moreover, users seemed to prefer to spend more time with it, perhaps without being conscious of the time they spent experimenting with various search strategies. On the other hand, when users were asked to be more precise and, especially, when nothing was retrieved, they seemed to grow discouraged. Lack of flexibility and lack of support, for example suggestions for alternative spelling or search terms, also seemed to further dishearten users from conducting searches.

13.2.3.3. Currency of Information

Currency of information was identified by interviewees as an important criterion for their choice of the Web as a first stop for information. In relation to this, one interviewee (Study3-Interviewee13) said “just [the fact that] information is found in a book does not mean it is up-to-date”. Another interviewee (Study3-Interviewee14) concurred, stating that “because a lot of my teaching is web-based, I start my search for information on the web. So [that] I get the most up-to-date information”. In addition, the latter interviewee mentioned the advantages of web-based information in getting oneself abreast of current developments and changes in a knowledge domain, usually accomplished through online subscriptions to updates and

newsletters. Whilst acknowledging the importance of printed books for high-level conceptual theories and concepts, which are relatively stable, it is notably limited in reflecting current developments within a field of study. Thus, one interviewee (Study3-Interviewee15) remarked: “In the area I am working on, books are too slowly paced”.

13.2.3.4. Access to Full-text

Whilst interviewees recognised the importance of printed books and the continuing role of the library as a physical space, they indicated their preferences for having access to both digital and printed versions of information, depending on their immediate requirements. Most interviewees highlighted the need for direct links to the full-text of books and articles, rather than mere bibliographic descriptions. Some interviewees expect their library to digitise as much of its collection as possible so that they are able to access and search with-in the electronic versions of these resources.

One interviewee (Study3-Interviewee2) suggested the need for creating a digital collection of books at the national level, so that users are able to access books that are not available in their local library. This however raises copyright issues that need to be addressed both at institutional and national levels. This suggestion is in line with that made by librarians, who have also highlighted the need for interoperation between libraries across universities nationally, so as to ease sharing of resources. Another interviewee (Study3-Interviewee14) mentioned their reliance on digital copies of books. The interviewee indicated their preference for buying digital versions of books, rather than printed ones.

13.2.3.5. About-ness Metadata and Searching inside Books

Interviewees indicated the importance of metadata about the content (about-ness) of an information object rather than the physical description (such as its format and location) and bibliographical details (Title, Author and Year of Publication). Interviewees also recognised that librarians, unless they have knowledge of the specific subject domain they are working in, may not necessarily be able to adequately describe the content inside an information object. This challenge, as the interviewees noted, can be addressed by involving users to provide about-ness metadata. In addition, some information objects such as “artists’ books” (an illustrated and handmade pieces of visual artefacts prepared as books) require subjective expertise to describe them. In connection with this, one interviewee (Study3-Interviewee17) recalled their experience that they had to offer their professional support to describe such specialised forms of books. The unique genre of these books, the interviewee indicated, also necessitated special attention to the type of information displayed as metadata. The interviewee described the challenge as follows:

“Firstly, the librarians wouldn’t be able to do this. Secondly, I myself found it very difficult to do this. I had to do a lot of online searching to describe the books. In other cases, I had to look into the book and describe it. While a library does not necessarily display in its catalogue visual metadata, for artists’ books it is vital that people can see what the book looks like. So I always ended up having pictures of what the structure

of the book looks like, scanning the pages from the book, so that users can come to the collection and access through it” (Study3-Interviewee17).

For example for artists’ books, the interviewee (Study3-Interviewee17) identified visual metadata, and ‘Literary and Visual Analysis’ was considered an essential piece of information. The interviewee indicated the importance of collaboration between librarians, authors, publishers and users in describing such specialised books. Since libraries cannot possibly hire subject knowledge experts for their various collections, collaboration through social media (Web 2.0) approaches is suggested as a feasible possibility, thus indicating the importance of socially-constructed metadata. The interviewee believed that collaborative approaches and the use of emerging technological platforms such as blogs can be best suited to make sense of these collections.

Interviewees indicated their preference for searching inside books. In this regard, they mentioned the importance of electronic versions of books. Most interviewees were unanimous in identifying abstracts, table of contents, book chapters and book snippets as useful pieces of information in determining the relevancy of information objects. Interviewees suggested the need for such metadata in addition to bibliographic details such as Title, Author, Publisher, Keywords and Year of Publication. It is important to note that this latter metadata is a priori metadata, which comes from the library or publisher. The interviewees cited their experience in benefiting from book chapters made available through commercial search engines and online book sellers (e.g. Amazon and Google Books). The interviewees also asserted that they saw no reason why such features cannot be incorporated in library catalogues. In connection with the benefits of such metadata, one interviewee (Study3-Interviewee1) noted that “I always like to go to the table of contents first, because I can then decide to go out and get the book in the library”.

Interviewees criticised the absence of book previews and table of contents in their current library catalogues. One interviewee remarked that “I tried to search for books in the library website, you have very little information. For example, it does not have summary [abstract] on what the book is about. Title is not enough. Before scanning the whole book, it would be good to have more information about it” (Study3-Interviewee5). Another interviewee (Study3-Interviewee6) indicated that getting book previews saves them time by offering them the possibility to scan through the contents before they borrow it. Another interviewee (Study3-Interviewee8) also identified their preference of Google Scholar and Amazon for their ability to provide previews and snippets of books. In connection with this, most interviewees maintained that richer metadata is important to judge the relevancy of information objects, thus suggesting their preference to see such features in future library search and discovery interfaces.

13.3. Saturating the Core Category of Metadata Filtering

The interviewees discussed their preferences and requirements with regard to convenience of access (13.2.3.1), instantaneous access (13.2.3.2), up-to-date-ness (13.2.3.3), access to full-text (13.2.3.4) and about-ness metadata (13.2.3.5). These features, they indicated, need to be considered when library systems are designed. Thus, these concepts are found important to

confirm the importance of conceptualising current metadata interfaces. Thus, in the future these requirements should be included in the principle of metadata filtering. The main purpose of Study Three was to determine whether the perspectives of library users saturates, extends, or confirms the results of Study Two, mainly on the core category of metadata filtering. The focused coding of interview data with library users resulted in the identification of concepts, mainly **traversal across information sources** (13.3.1), **sifting through relevant information** (13.3.2), and **triangulating information from disparate sources** (13.3.3) (see Figure 13.1).

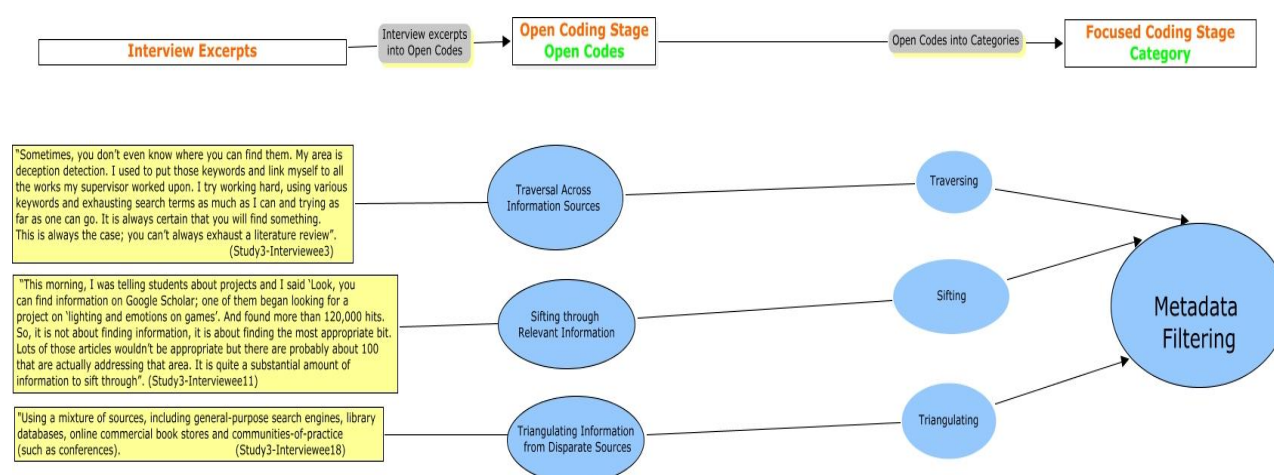


Figure 13.1: Saturating Metadata Filtering

13.3.1. Traversal Across Information Sources

Interview data reveals that users like traversing across multiple sources of information, rather than being confined to a single, or even just a few. They cross over from one information source to another, going, say, from Google to specialised library databases. Many interviewees admitted starting searches for general information on Google, when, for example, checking for facts, which usually leads them to Wikipedia or related sources. They rely on peers and professional networks, as well as recommendations from others, for identifying relevant information sources. They also consult both print and digital information sources, all of which are available in multiple formats and on diverse platforms. During traversal, users mostly conduct generic searches in domains, which they want to investigate in an exploratory search manner.

These searches can be made by users not well-versed in a given domain. This means that they are less likely to know the exact title, author or date of an article or a book. They are also not familiar with the information provider, such as a publisher, and, hence, conduct broader searches for a given topic of interest. They also employ generic and natural language search terms or keywords. For these kinds of searches, users seem to have found Google as the most user-friendly and simple search engine as it permits natural language keyword searches, while simultaneously offering search term suggestions. Google Scholar also permits searches with generic keywords and again offers suggestions. On the other hand, specialised databases, as users indicated, mostly require their users to be precise. For example, library

catalogues require users to be exact and presume that users come with a specific subject term, author name or title. Users, however, are aware of differences between academic library search catalogues and general-purpose ones.

Exploratory searches can also be conducted by experienced users who want to keep abreast of developments in their fields of interest. One interviewee noted:

“Sometimes, you don’t even know where you can find them. My area is deception detection. I used to put those keywords and link myself to all the works my supervisor worked upon. I try working hard, using various keywords and exhausting search terms as much as I can and trying as far as one can go. It is always certain that you will find something. This is always the case; you can’t always exhaust a literature review” (Study3-Interviewee3).

In contrast to exploratory searches, for focused searches users already know the title, author or combination of keywords to be employed for finding a specific article or topic. Its difference from an exploratory search is that the user knows what to expect. Focused searches are mostly conducted by researchers who have already developed experience in a particular knowledge domain. This is often referred to as an information object search. Users know specific databases journals to be consulted when looking for such information. For example, if a user approximately knows the title of an article, they may search until they find an exact match, thus everything else may be considered as a mere distraction. However, the aim of such an exercise is access to the information object. The success of a focused search can be measured by the match between what was sought for and what has been retrieved.

Through experience, and over time, users seemed to develop skills that enable them to coin relevant keywords when formulating a search. For example, one interviewee (Study3-Interviewee9) observed “once you get conversant with the terminologies used in your research area, it becomes easier to search for and find relevant resources”. The interviewee highlighted the importance of employing specific and pertinent keywords, if one is to gain efficiency during searches. Most interviewees were aware of the difference in using broad versus specific keywords, as the former is likely to result in too many retrievals, some of which are quite irrelevant. One interviewee (Study3-Interviewee14) remarked thus, “you don’t want it to be a case of being too broad, and, then, every search returns every single book, resource or whatever”. Users also try as varied keyword combinations as possible in their attempts to retrieve pertinent sets. The speed with which retrievals are displayed is crucial for users, when trying several search formulations. In connection with this, one interviewee (Study3-Interviewee22) described their own experience, stating “I start from a broad set of keywords, which give me a broad set of results. Then I narrow it down to a specific sub-topic. For example, I start a search using ‘petroleum pricing’, and then narrow it down to ‘petroleum pricing in Africa,’ and then to ‘petroleum pricing in West Africa’.

In addition to employing various search formulations, users may also browse library shelves physically or taxonomies of electronic databases, all with the aim of finding something pertinent. Users may also subscribe to journal feeds and alerts. One interviewee (Study3-

Interviewee20) for example stated that “The way I update my disciplinary specialisation is usually through technology, ones that lets me know when new journal articles are published, calls for papers are made, and conferences are held”.

Whilst all interviewees seem to share consensus on the need for traversal across various information sources, they nonetheless noted the importance of evaluation and critique. One interviewee (Study3-Interviewee15) believed that any information found on the web should be taken with a “pinch of salt”. The same metaphor was also employed by another interviewee (Study3-Interviewee14) who believed that evaluation and critiquing skills should be life-long learning experiences. In relation to information evaluation, almost all interviewees acknowledged the need for objectively ascertaining the quality of postings and facts in such sources as Wikipedia, where anyone can edit and change them. The importance of traversing across information sources calls for consideration of such users’ behaviours in the design and implementation of OPAC and discovery interfaces.

13.3.2. Sifting through Relevant Information

The ability to sift for relevant information was identified as an important skill. As the interviewees suggested, the volume of information grows, at an increasing pace and, consequently, as search engines fetch retrieval sets that are often too many, it requires appropriate user know-how and evaluative judgement to quickly sift through the retrieval in order to separate the pertinent from the irrelevant.

“This morning, I was telling students about projects and I said ‘Look, you can find information on Google Scholar; one of them began looking for a project on ‘lighting and emotions on games’. And found more than 120,000 hits. So, it is not about finding information, it is about finding the most appropriate bit. Lots of those articles wouldn’t be appropriate but there are probably about 100 that are actually addressing that area. It is quite a substantial amount of information to sift through” (Study3-Interviewee11).

Elaborating further, the interviewee stated, “I suppose in our information-rich age, there is lots of information out there and it is a matter of selecting what is most useful. It is about selecting the most appropriate bits from the whole cloud in that particular subject area.” It is indicated that once users get initial results, they are more likely to refine searches. Once they have a general idea of what they are looking for, they learn from the search experience.

Sifting through diverse information sources, cross-checking among information sources and evaluating the credibility of the information has become an important skill. Regarding the authority of information sources, one interviewee (Study3-Interviewee15) was of the opinion that “the publisher is irrelevant” for judging the quality of the information. According to them, any possible source of information is valid. Furthermore, the process of information evaluation has shifted from one that relied on a publisher’s filtering mechanisms to ones devised by the individual user. Hence, the ability to evaluate and utilise information has become ever more crucial. This is especially true for lecturers, who seem to be self-assured

that they have acquired the requisite experience and skills for the task. For example, their use of Wikipedia is predicated on their ability to weigh the credibility of the information.

Metadata filtering should therefore take into account such amount and diversity of information sources and the time requirements of users to sift through such volumes of information sources.

13.3.3. Triangulating Information from Disparate Sources

Whilst traversing across diverse sources, interviewees stated that they triangulated the information thus acquired, employing the most appropriate sources when conducting research, teaching or learning. The choice of sources, interviewees indicated, mostly depends on the task to be accomplished: writing an essay, finding factual information, preparing a lecture or conducting research. One interviewee (Study3-Interviewee18), for example, indicated that they used a mixture of sources, including general-purpose search engines, library databases, online commercial book stores and communities-of-practice (such as conferences). The extent of triangulation also varies, depending on the type of information required. Searches for factual or generic information, some interviewees noted, can be answered with general-purpose search engines, such as Google or even Wikipedia. For academic and research purposes, especially in specialised topics, on the other hand, users try to use as many sources as possible and use various databases, e-journals, peer/supervisor recommendations, print and electronic books and a plethora of other sources, with the aim of (almost) exhausting all available sources.

In connection with this, interviewees expressed their wide experience in mixing-and-matching information from diverse sources, which they have acquired while traversing the Web and related information services. One interviewee (Study3-Interviewee15) recalled, with amusement, how they managed accessing information sources prior to the advent of the Web, with sole dependence on the library. The interviewee further recollected that, in a matter of a few years, unprecedented technological changes had happened, all influencing the way information is accessed and utilised.

In accordance with the demands of their professional roles, researchers and lecturers are more likely to combine deep searching and browsing approaches, traversal across information sources and triangulation from disparate sources, when meeting their information requirements, as it is highly unlikely that a single search strategy or a single source will adequately satisfy their needs. It should be evident that seamless linking among disparate sources of information is of critical importance, whereby it is imperative that libraries pursue the goal of seamlessly linking diverse information sources. As pointed out by most interviewees, no single library could satisfy all their needs at all times, which brings to the fore the need for libraries to cross-link, not only amongst themselves, but with external information sources too, in order to facilitate seamless navigation among as many sources as possible. Users believed that such linking would significantly improve the probability of their discovering relevant sources.

13.3.4. Socially-Constructed Metadata

Interviewees variously indicated the importance of involving library users in metadata in the form of tagging, reviews, ratings and recommendations (see Figure 13.2). Such socially-constructed metadata, interviewees agreed, would add value to library services, especially in identifying and choosing relevant information sources, thus saving users time and effort. In addition, in the context of universities, interviewees saw the benefit of user-generated metadata to encourage the use of reference lists and also to gauge the usage of books recommended by lecturers. Rating systems are considered one of the ways to achieve this.

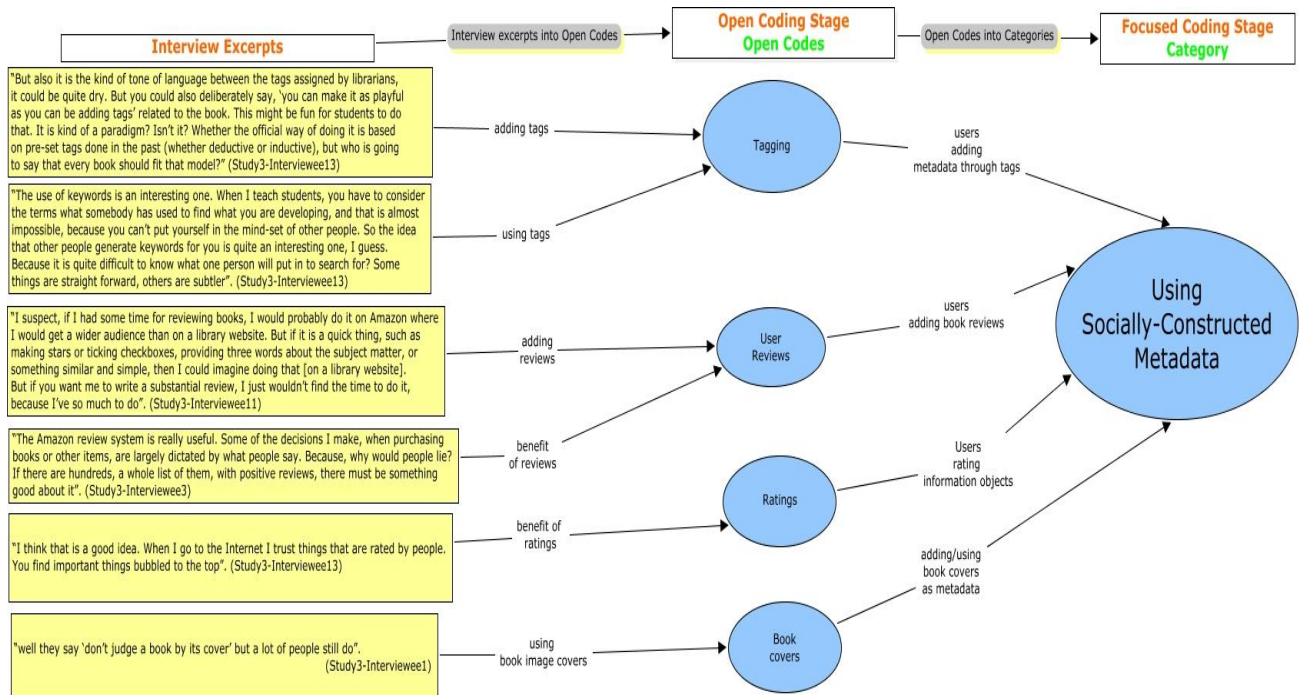


Figure 13.2: Using Socially-Constructed Metadata

Some interviewees are well aware of socially-constructed metadata applications such as GoodReads, Amazon and e-Bay. One interviewee (Study3-Interviewee11) for example indicated:

“I think that the way Amazon works is that there is a whole crowd sourcing going on, where you get people writing about books in the reviews, and often quite detailed ones. That works really well because that way you get different perspectives. So that could be very helpful. I think if university libraries want to put more information on books online, that wouldn’t be a bad way to do it, rather than one person doing it or an academic writing a description of the book or even a description from the publisher. But actually getting people who have read and used that book to comment (review) would be something useful. Or maybe you don’t need to do that and instead just put a link straight to the reviews on Amazon or Google Books. And connect people to these resources”.

Judging the relevancy of an information object such as a book or a journal article, the interviewees noted, is one of the most important and often time consuming component of their information skills. One interviewee (Study3-Interviewee6) indicated the importance of knowing other readers' perspectives through reviews. The interviewee further stated that "even before you read a book, you can have some [prior] knowledge about it. What are the thoughts of people who have read the book? What are the reviews? You don't want to end up studying the wrong book. That may be wasting a lot of time".

The next sub sections discuss some of the applications which the interviewees indicated as important and exemplify the Core Category of metadata enriching (see Chapter Nine). These include:

- Tagging (13.3.4.1)
- User reviews (13.3.4.2)
- Ratings (13.3.4.3)
- User recommendations (13.3.4.4)
- Book covers as metadata (13.3.4.5)

13.3.4.1. Tagging

The interviewees considered tagging of information sources as a helpful tool in their search and discovery experiences. One interviewee (Study3-Interviewee13) considered tagging as a good tool to incorporate user terminologies in library metadata.

"But also it is the kind of tone of language between the tags assigned by librarians, it could be quite dry. But you could also deliberately say, 'you can make it as playful as you can be adding tags' related to the book. This might be fun for students to do that. It is kind of a paradigm? Isn't it? Whether the official way of doing it is based on pre-set tags done in the past (whether deductive or inductive), but who is going to say that every book should fit that model?"

The interviewee believed that the perspectives (interpretations) of students with regard to books and other information sources (post-hoc) metadata is as important as the expert (librarian) created metadata. The interviewee, however, also recognise the technical and social challenges associated with implementing this approach to metadata. Another interviewee (Study3-Interviewee14) added that:

"The use of keywords is an interesting one. When I teach students, you have to consider the terms that somebody has used to find what you are developing, and that is almost impossible, because you can't put yourself in the mind-set of other people. So the idea that other people generate keywords for you is quite an interesting one, I guess. Because it is quite difficult to know what one person will put in to search for? Some things are straight forward, others are subtler".

Note that the findings of Study Two with LIS professionals (librarians, researchers, lecturers

and metadata consultants) also identified tagging as one of the primary instances of socially-constructed metadata approaches (8.4.1.1).

13.3.4.2. User Reviews

Most interviewees referred to their experiences in using product reviews during on-line shopping. They also acknowledged familiarity with user books reviews. It was also mentioned that the rating of these reviews as well as incorporating provenance information, such as the source and context of each review, would bolster users' confidence in taking advantage of them. User reviews, as some interviewees remarked, should be used with care. But almost all criticised the absence of provision for incorporating user reviews on library websites.

Whilst most interviewees are cautious with the multi-faceted implications of socially-constructed metadata including the motivation to get students to write relevant and sensible reviews, quality of the metadata and policing issues, many interviewees saw the benefit of user generated metadata in the university context. With regard to book reviews, one interviewee (Study3-Interviewee11) indicated that "Because of not having the time to go and read all the books and make judgements about them, you rely on other people telling you. If I had time and to go and read all those books that would be great. But I don't. So I have to go and find reviews". The interviewee also believed students would benefit from user generated metadata such as user reviews and recommendations.

Another interviewee (Study3-Interviewee14) also indicated the importance of socially-constructed metadata especially tagging and ratings. However, when it comes to book reviews, the interviewee was cautious and believed that as a professional they want to see well-informed critiques rather than a ten-second snapshot of reviews which may mislead other users. The interviewee commented at length stating:

"A review which may be seen by a lot of other people on the web, with your name in it, doing it with a bullet point isn't really doing yourself or the book justice. So, I think that is a very interesting idea, and indeed I used reviews on Amazon for books, and see what other people said, but you can also see whether the person who wrote the review is an expert or a lay person? That is kind of handy to know without having to look at more of what they have written, is the review educational? or is the resource a valid resource? Some people may just want to slam out some words because they can, because obviously now with social media and self-publishing on the Internet, I think there are a lot of people who just want to get their words out, regardless" (Study3-Interviewee14).

With regard to user reviews, one interviewee stated, "The Amazon review system is really useful. Some of the decisions I make, when purchasing books or other items, are largely dictated by what people say. Because, why would people lie? If there are hundreds, a whole list of them, with positive reviews, there must be something good about it" (Study3-Interviewee3). Another interviewee expressed their preference for posting their reviews on widely accessible and popular websites, such as Amazon.com, rather than on a library one,

adding, “I suspect, if I had some time for reviewing books, I would probably do it on Amazon where I would get a wider audience than on a library website. But if it is a quick thing, such as making stars or ticking checkboxes, providing three words about the subject matter, or something similar and simple, then I could imagine doing that [on a library website]. But if you want me to write a substantial review, I just wouldn’t find the time to do it, because I’ve so much to do” (Study3-Interviewee11). This interviewee further recommended that, instead of duplicating efforts, the library community should try to harness such user reviews that are available on commercial or other book review websites. Overall, notwithstanding misgivings in relation to issues associated with motivation, linking to external websites, quality of reviews, and the possibility of abuse (in the form of misleading or misinforming users), most interviewees considered reviews as important examples of socially-constructed metadata (see also 8.4.5.7). They conserve precious users’ times by enabling filtering for relevant information sources while simultaneously providing a wide range of user perspectives on a given information object.

13.3.4.3. Ratings

Ratings, as some interviewees believed, would help inform users about the relevance or popularity of information objects. In connection with this, one interviewee (Study3-Interviewee13) noted that “I think that is a good idea. When I go to the Internet I trust things that are rated by people. You find important things bubbled to the top”. In comparison to other socially-constructed metadata approaches such as tagging, reviews or recommendations, as some interviewees indicated (Study3-Interviewee5; Study3-Interviewee6; Study3-Interviewee2), rating is relatively light as it does not require much user effort. Some users preferred a Facebook approach to rating such as ‘Likes’, whilst others preferred Amazon’s approach to rating using a five-star rating systems (compare also with discussions in 8.4.1.3).

13.3.4.4. User Recommendations

User recommendations, as revealed by interview data, play an important role in the discovery of information objects by users. All interviewees mentioned instances where they have used recommendations by peers and colleagues. Interviewees recalled instances of getting invaluable information at conferences (Study3-Interviewee9; Study3-Interviewee4). It was also mentioned that researchers in project or research teams also share recommendations amongst themselves. Another interviewee recalled receiving suggestions for books from colleagues and friends, which was particularly pertinent for their course, and further stated their belief that it is up-to the user to go through the metadata given in the recommendations (Study3-Interviewee16).

Other interviewees also mentioned social media as an important avenue for sending recommendations. Regarding this, one interviewee (Study3-Interviewee4) stated their habit of keeping their students abreast of current developments, employing Twitter feeds. However, the interviewee noted that such practices of using social media works best when many students are involved. To quote them verbatim, the interviewee asserted, “On a library

website, I think social media would be a lot better for students who are doing general courses, rather than those with specialised research interests, such as a PhD” (Study3-Interviewee4). The said interviewee also noted the importance of avoiding spoon-feeding students, arguing that students should develop information seeking and evaluation skills; pointing out the benefits of finding information in one’s own ways, including learning from failures in finding relevant information. However, interviewees admitted that not all recommendations are accurate or relevant. Some interviewees, for example, alluded to problems in some of the recommendations from Amazon (Study3-Interviewee8; Study3-Interviewee12; Study3-Interviewee14), thus highlighting the importance of accurate value judgements on the part of the reader. See also the perspectives of LIS professionals on use and system recommendations (8.4.1.4).

13.3.4.5. Book Covers as Metadata

Interviewees maintained that the cover of a book is an informative visual metadata. It is also noted that it adds an aesthetic value to the display. One interviewee (Study3-Interviewee1) remarked that “well they say ‘don’t judge a book by its cover’ but a lot of people still do”. In the context of the physical library, book covers speed up identifying and locating a book on a shelf. The interviewees indicated that they sometimes waste a lot of time wandering around shelves and such metadata would help them save time. One interviewee (Study3-Interviewee12) also mentioned their use of book covers for their reference lists. Another interviewee (Study3-Interviewee16) also noted that the cover of a book creates an impression about the book which sometimes affects decisions to choose the book. Overall, whilst it is one piece of a priori metadata, book covers play an important role to enhance the findability and discoverability of books in libraries.

13.4. Conclusions

The analysis of Study Three confirms the importance of metadata filtering that emerged in Study Two. In Chapter Twelve, the importance of providing re-configurable, user-led, multiple metadata views (displays) is suggested (see 12.2). In Chapter Thirteen, users indicated the importance of convenience of access (13.2.3.1), instantaneous access (13.2.3.2) and triangulating information from various sources (13.3.3), all indirectly requiring the re-conceptualisation of current (expert-led) single metadata displays.

In addition to improving expert-based (a priori) metadata, interviewees indicated the importance of socially-constructed (post-hoc) metadata (13.3.4). In relation to post-hoc metadata, users think reviews (13.3.4.2) are important to get support from credible sources such as peers or lecturers, about the relevancy of a specific information source. Rating (13.3.4.3) the reviews was mentioned as a way to ensure the credibility of reviews and socially-constructed metadata. Most users mention that such socially-constructed metadata should be used with a “pinch of salt”, indicating the importance of checking the provenance of user-created metadata (see also 8.4.5.8). Some users fear that such metadata may be commercially motivated or out of intended malice, thus users should develop evaluation skills.

Users indicated their preference for richer metadata displays, including tables of contents, abstracts, reviews and book covers (13.3.4.5). In addition, users preferred simple interfaces, but at the same time they wanted to be provided with richer metadata displays. Thus, complexity of data should be provided within a simple interface. The data analysis in Chapter Twelve (metadata filtering) also suggested the need for hiding complexity. The analysis in Study Three confirms the findings in Study Two where librarians also indicated the importance of re-orienting current metadata interfaces (see Figure 13.3).

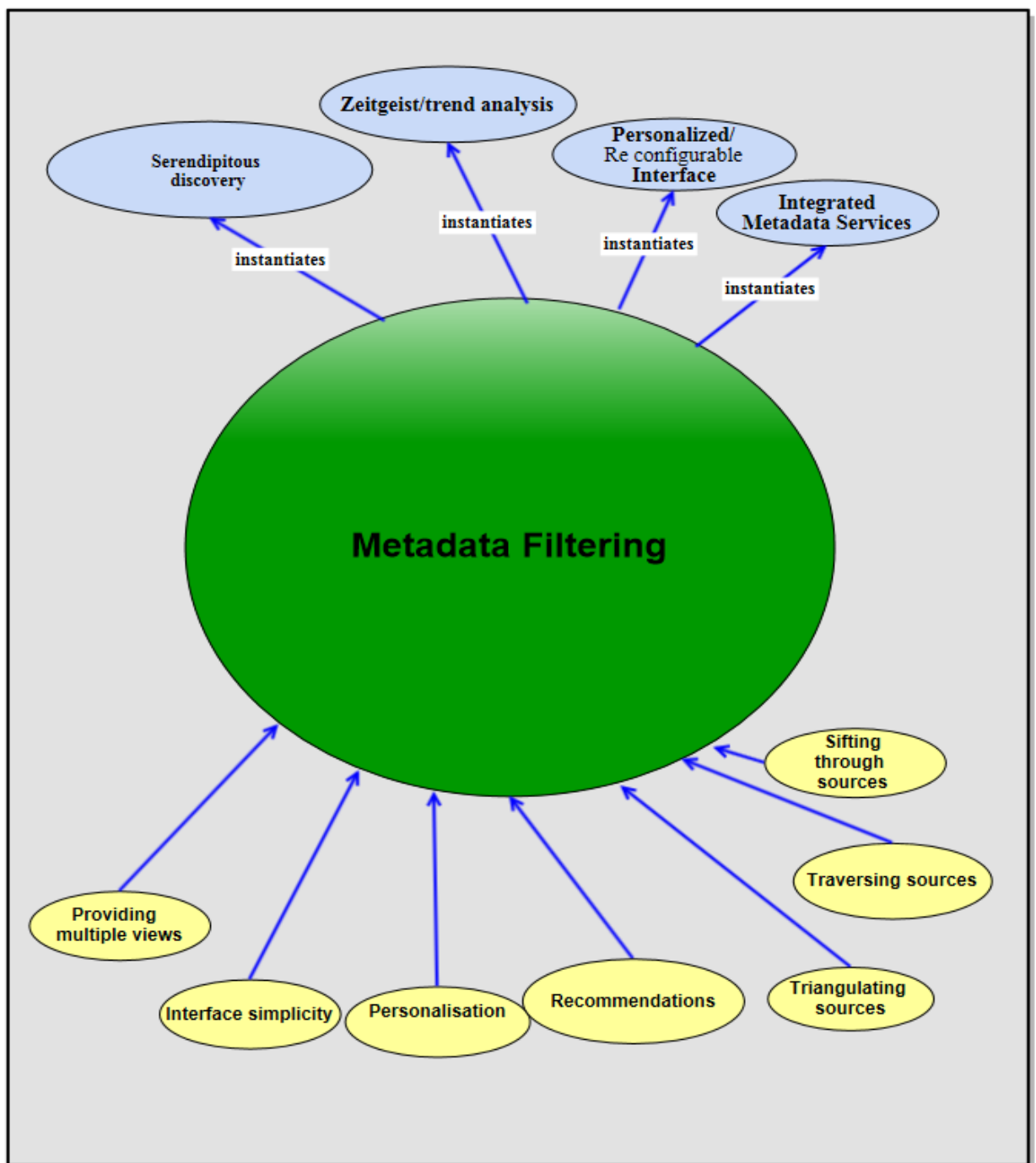


Figure 13.3: The Core Category of Metadata Filtering

Confirming the results of Study Two, the data analysis in Study Three showed the importance of providing multiple views (12.2) with easy to use and easily reconfigurable displays and providing library services catered towards the needs of users by using personal profiles and use (search) behaviours (12.4). In addition, the results of Study Three, indicated that library users tended to prefer convenience (13.2.3.1), immediacy (13.2.3.2), currency (up-to-date-ness) (13.2.3.3) and access to full-text (13.2.3.4) as primary reasons for choosing current search engines, such as Google as their first stop for information (see 13.2.3). The interviewees also indicated their search and discovery behaviours are characterised by traversing, sifting and triangulating information sources. The incorporation of these issues in the process of metadata filtering is considered an important factor in addressing the needs and requirements of library users. Overall, the findings in Study Three saturate the Core Category of metadata filtering (Chapter Twelve). Informed by the categories identified in chapters Twelve and Thirteen, metadata filtering emerges as an important metadata principle.

Chapter Fourteen: The Emergence of a Theory of Metadata Enriching and Filtering

14.1. Chapter Overview

Resulting from the suggested benefits of a Mixed Metadata Approach (9.1), where the case was made that a priori metadata and post-hoc metadata are considered as complementary to each other, further iterative data analysis resulted in the emergence of four overarching principles. These were metadata enriching (Chapter Nine), linking (Chapter Ten), openness (Chapter Eleven) and filtering (Chapter Twelve & Thirteen). This chapter describes and discusses the interdependence between the principles of enriching, linking, openness and filtering. Finally, the main theory, the Theory of Metadata Enriching and Filtering, is discussed.

14.2. Integrating the Four Principles

The principle of metadata enriching shows a continuous process of adding, augmenting and improving expert-created (a priori) metadata with user-created (post-hoc) metadata, whereby the latter provides the diversity and breadth of interpretations of information objects. A priori metadata serves as a hook to present basic structured standards-based metadata to users. Once users get access to information objects using a priori metadata, they can start enhancing it using their own terminologies, interpretations or descriptions. Enriching is thus characterised by a constant flux. Enriching can be contrasted with the current metadata principle of sufficiency and necessity (2.6), which focuses on creating simple metadata in a ‘complete state’.

Whilst it is technically possible that libraries can implement the principle of enriching within their own institution, however it can only be effectively and efficiently implemented if another principle, the principle of metadata linking, is embraced and implemented. The principle of linking enables libraries to continually enrich their existing metadata with metadata that resides outside their boundaries. Metadata that is linked, both with internal and external data sources, results in interlinked metadata, thus offering users the ability to seamlessly navigate between disparate information objects. The principle of metadata linking ensures that metadata values are granularly structured, uniquely and persistently identified, and interlinked, thus, bringing together disparate metadata sources. However, in order to realise metadata linking, the metadata that is to be linked-to must be made openly accessible. The principle of metadata openness states that institutions communally benefit from making their metadata available in an open, re-usable and re-combinable format. Whilst there can be various degrees of openness, the highest degree of openness where metadata is open, machine process-able and interlinked provides the greatest benefit, in terms of opportunities for re-use.

The principles of enriching, linking and openness are interdependent, however, the latter two can be considered as subsumed under the former. For the two principles to be effectively utilised, a higher goal of the principle of enriching should be first put in place. As a

consequence of enriching being a broader goal, the principles of linking and openness are considered as required components of metadata enriching (Figure 14.1).

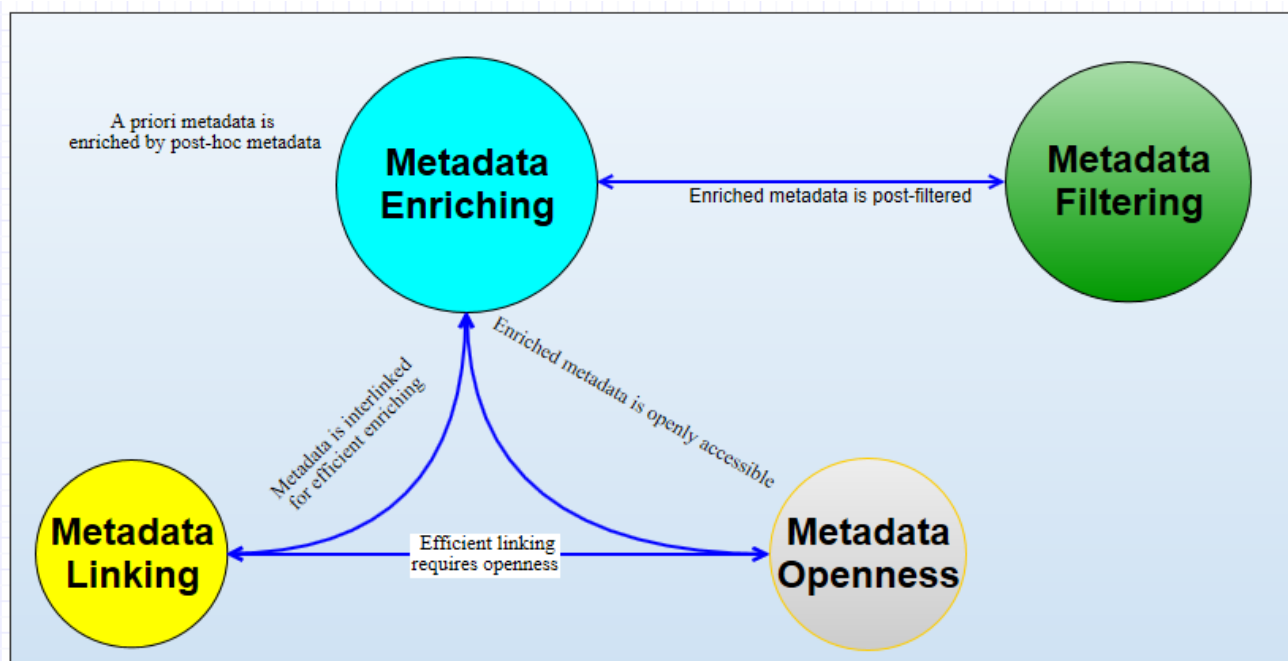


Figure 14.1: Principles of Metadata Enriching, Linking, Openness and Filtering

The interdependence and subsequent integration of the three principles resulted in the overarching principle of enriching. The principle of metadata enriching is bound to result in a great volume of metadata. Thus, without appropriate filtering, metadata enriching would become more of a problem than a solution. Unless properly filtered, the sheer volume of metadata presented to users on discovery interfaces (such as OPACs), may hamper the findability of information objects, as users find navigating through the retrieved mass difficult or time consuming. This is where the principle of filtering plays a crucial role. Thus, the overarching principles of enriching and filtering emerged, resulting in the emergence of the Theory of Metadata Enriching and Filtering.

14.3. The Theory of Metadata Enriching and Filtering

The main tenets of the Theory of Metadata Enriching and Filtering include the emergence of a mixed metadata paradigm (interdependence of a priori and post-hoc metadata), the consideration of users as proactive co-creators, a shift in the role of librarians from metadata content creators to metadata systems architects, the re-conceptualisation of metadata quality and the separation of metadata content from its presentation (interface). The Theory of Metadata Enriching and Filtering provides an overarching framework for the concepts and principles developed in this thesis. The theory therefore elegantly abstracts and represents a complex set of concepts and principles.

In summary, since the principles of metadata linking and openness are subsumed under the principle of enriching, the emergent theory can be refined to two major principles:

1. **Metadata enriching** which states that a priori metadata that is instantiated and granularly structured by library metadata experts is continually enriched through socially-constructed (post-hoc) metadata, whereby users are pro-actively engaged in co-creating metadata. The principle also states that metadata that is enriched is also contextually and semantically **linked** and **openly accessible**; and
2. **Metadata filtering** which states that metadata resulting from implementing the principle of enriching should be displayed for users in line with their needs and convenience.

14.3.1. Separation of Metadata Content (Enriching) and Interface (Filtering)

The theory emerged from this research makes a vital distinction between enriching and filtering. Whilst enriching solely deals with metadata content, filtering addresses all issues associated with its presentation (interface). The former is about complexity and the latter is about simplicity and efficiency. Simplicity is a usability issue, thus should not be conflated with enriching of metadata content. In the emerging metadata paradigm, the challenges that arise in presenting the sheer volume of metadata are addressed by the principle of filtering. Current principles mistakenly assume to achieve simple presentation of metadata (i.e., metadata simplicity) by reducing the number of metadata fields by applying what is called Occam's razor (see 2.6 Principle of Sufficiency and Necessity). However, the emerging Theory of Enriching and Filtering espouses to collect as much metadata as possible using a mixed metadata approach, and apply post-hoc user-led filtering. The theory suggests that metadata that is enriched should be presented to the user in a simple, usable, sensible and meaningful manner. Whilst the principle of enriching results in a potential abundance of metadata, the principle of filtering is used to simplify its presentation by enabling a user-centred/focused/led design.

14.3.2. Separation of About-ness from Medium

The Theory of Enriching and Filtering places primary importance on describing and interlinking about-ness rather than the physical mediums (carriers) of information objects. Thus, it focuses on what is contained within the medium, such as books and journals. Some of the current metadata limitations arise due, first of all, to the fact that current metadata standards are mainly concerned with describing the physical characteristics of information objects (medium) rather than the content found in them (message) (Day, 1997 as cited by Wright, 2007, p. 86). Even though metadata that describes the medium and its location on shelves or the computer network is important to ensure the accessibility, provenance and authenticity of information objects, metadata that deals with the content of an information object is crucial for making sense of its essence and, hence for adding value to its use. As enriching the content requires additional knowledge of the subject matter, the role of users as metadata co-creators (8.4.2) is essential.

It is argued that the emerging Theory of Enriching and Filtering frees information from their medium. The users' preponderance to traverse across information sources (mediums) (13.3.1), wanting to triangulate information from disparate sources (13.3.3), necessitates that

the focus be placed on interlinking ideas contained in various databases, books and journals. The freedom of information from its constraining medium through seamless interlinking should help users to connect to inter and multi-disciplinary domains of knowledge.

14.3.3. Enriching and Filtering as a Non-Deterministic Process

Whilst at present standards-based metadata approaches such as taxonomies and controlled vocabularies attempt to predetermine (anticipate) terminologies a priori, it is nonetheless a difficult and untenable exercise. In light of this, the Theory of Enriching and Filtering recognises the importance of capturing the usage of terminologies post-hoc and continually improving and enhancing metadata through collective metadata intelligence (8.4.5).

Thus, enriching is considered a non-deterministic process. It does not aim to anticipate future metadata needs based on a priori metadata. Instead by using a priori metadata as a structure to get access to post-hoc metadata, it learns from the collective metadata intelligence. Through a mixed metadata approach, the principle of enriching enables libraries to identify the zeitgeist and resource usage patterns of library users. It helps libraries to continually re-structure and enhance metadata after-the-fact (post-hoc) rather than a priori. Hence, librarians can continually improve the structure (7.4.1), granularity (7.4.2), provenance (7.4.3) and interoperability of metadata (7.4.6). A non-deterministic view of enriching and filtering prefers to record user preferences post-hoc rather than attempting to anticipate users' needs. In addition, the principle of enriching takes the view that terminologies change through time. It takes an interpretivist rather than an objectivist ontological point of view.

14.3.4. From User-Centred to User-Driven Metadata Enriching and Filtering

Librarians have grappled with the question of objectively ascertaining users' requirements through surveys, interviews and feedback forms. However, the answers thereby obtained can only make library metadata user-centred but not user-driven, the latter being a more desirable outcome. Only direct involvement, in the creation of metadata on the part of users, can assure that metadata is indeed truly user-driven. User-driven metadata enriching can empower users to be proactive creators, collaborators and partners. In the new metadata paradigm, users co-own the metadata. They can participate not only in the co-creation process, but also in its management and curation. Since the main objectives for affixing metadata to information objects is to enable its discoverability, metadata should be judged by its relevance in meeting this objective, rather than the way it depicts an information object in an objective and ontological manner. Terminologies change over time, hence, the metadata should be able to be cognisant of and reflect such changes.

14.3.5. Enriching as a Continuous Process

In contrast to creating a priori metadata in a 'complete state', enriching is a continuous process of adding, enhancing and improving metadata content. Enriching is thus a never ending but evolving process. Benefiting from a network effect (8.4.5.5) where each contribution slowly aggregates (8.4.5.3), enriching aims to collect as much metadata as possible, and thus facilitating metadata diversity (8.4.3).

All interpretations are valid; thus any user can add anything at any time, provided that such metadata assertions are not malicious. As new users join the network, it is possible that the metadata becomes further enriched (note: enriching is a continuous process and it will never be enriched completely). However, with new conversations taking place regarding a given information object, its interpretation has the possibility to continuously evolve (change) rather than simply refine (consistency).

14.3.6. Metadata Diversity better Conforming to Users' Needs

Metadata diversity (8.4.3) implies the inclusion of a multitude of potentially conflicting metadata ascribed to information objects by users. Metadata diversity increases the likelihood of conforming to the multitude of perspectives and interpretations of various groups of potential users. In relation to idiosyncratic (personal) metadata entries, it is important to recognise that a given metadata entry that might be considered trivial for a general user might be important for the one who created it, since it is likely that the latter will search with those keywords. Such idiosyncrasy can be managed by providing personalised presentation, which can be managed through appropriate metadata filtering (Chapter Twelve). The new metadata paradigm may thus permit the inclusion of metadata descriptions (interpretations) of information objects that may seem in opposition. In such instances, it is important that the diversity of the various interpretations is maintained. Nevertheless, such a metadata paradigm should not include random entries; instead, it should cater towards semantic and meaningful metadata whilst at the same time maintaining the diversity of interpretations.

14.3.7. Ubiquitous Linking

At present the links to metadata on an OPAC display can only go up to a certain extent and soon reach a dead end, when a data item is not linked any further. For example, current metadata interfaces allow a user to search for a specific book, click on the author's name and see their publications, and, maybe find related/similar books. They, usually, do not offer a seamless link to the author's biographical page or from there to another page, such as one on Wikipedia and Google Books, that cites him, and back to the library listing.

The Theory of Enriching and Filtering posits that metadata that is enriched with links would give endless possibilities to explore and discover information objects. Potentially, every metadata value can be linked in a similar way as words and phrases in a dictionary can be linked, thus users can select any word and retrieve the meaning of it. Likewise, users can select any metadata link retrieving information objects associated to that metadata value. For links to be ubiquitous, metadata openness is essential so that linking to external sources can be effectively implemented. Metadata openness has far reaching consequences for the way metadata in libraries is created, accessed, shared and re-combined. If library metadata needs to be made re-usable and shareable, it should be represented in open formats and should be made available in flexible licensing schemes that allow not merely free use, but predominately adaptations and commercial uses. Furthermore, the linking is designed in such a way that it does not get in the way of users; rather it is seamlessly integrated.

The Theory of Enriching and Filtering provides an integrated approach to metadata linking. Whilst at present metadata linking is considered in isolation to the processes of metadata creation and filtering, this theory argues that linking should be seamless and ubiquitous. At present users are forced to search various disparate databases within a given library, however, the Theory of Metadata Enriching and Filtering suggests that seamless metadata linking should be the way forward as it eases the navigation between various information sources without the user necessarily making extra effort to stop one database and start another. The need for seamless linking is especially apparent in the light of emerging user behaviours to traverse across information sources (13.3.1), sifting through relevant information (13.3.2) and the need to triangulate information from disparate sources (13.3.3). Potentially, every metadata value can be interlinked, thus users can select any metadata link retrieving information objects associated with that metadata value. Integrated and seamless linking is thus considered essential to achieve the overall goal of enriching.

14.3.8. ‘Useful’ rather than ‘Perfect’ Metadata

The theory suggests the importance of re-conceptualising the notion of metadata quality, stating that in a mixed metadata approach the issue of metadata quality should be weighed in light of the relevance of metadata to support findability and discoverability. Further, it is metadata’s usefulness for finding and discovering information objects, rather than on whether it is objectively accurate (truthful) or not, or on whether it is ‘good enough’, rather than on whether it is ‘perfect metadata’, that should be considered. The Theory of Enriching and Filtering considers all users as potential metadata creators (8.4.2). As a result, it is likely that good as well as erroneous (malicious) metadata may be added into the system (8.6). Due to the sheer volume of metadata entries, it is not possible for librarians to ensure the quality of metadata in terms of consistency. In this regard, it is important that the notion of metadata quality in libraries is re-conceptualised.

14.3.9. Post-Hoc Filtering

Whilst basic a priori metadata structures are put in place by metadata experts, the principle of filtering posits that users should be allowed to re-configure the layouts and preferences of display. In order to effectively deliver metadata to users and improve their experience of interface utilisation, filtering should be designed in such a way that users can make choices based on their requirements. Thus, the traditional types of filtering should also be re-conceptualised. Traditionally, filtering was done a priori by the service provider. In the light of the principle of metadata enriching, what is required is filtering that can be tailored by the users themselves, post-hoc (on-demand) filtering.

From the data analysis, the principle of metadata filtering emerged as one of the most important Core Categories (see Chapter Twelve). As part of the principle, important areas of concern that need to be addressed were indicated, including, providing multiple metadata views (12.2), interface simplicity (12.3), personalisation (users profiles) (12.4), support recommendations (12.5; 12.6), integrated access (12.7) and serendipitous discovery (12.8). The principle suggests the importance of re-conceptualising current metadata interfaces from

a single expert-led design to a multiple, re-configurable and user-driven design. The emergence of the principle of metadata filtering is also informed by users' preponderance to traverse across various types of information sources (13.3.1), the need to triangulate information from disparate sources (13.3.2), the importance of convenience of access (13.2.3.1) and instantaneous access to information (13.2.3.2).

Post-hoc filtering is primarily user-led and user-focused. Using collaborative filtering, recommendations and multiple re-configurable displays, the principle of metadata filtering better addresses users' needs and requirements than traditional (expert-led/single display) interfaces. The principle offers flexibility and facilitates serendipitous discovery of information resources (12.8).

14.4. Conclusion

By providing underpinning concepts and principles, the Theory of Metadata Enriching and Filtering emphasises the importance of addressing the 'why', instead of the 'how', to use particular technologies. Whilst at present libraries attempt to implement Web 2.0 technologies and instances of socially-constructed metadata approaches, such as tagging, these efforts are not always integrated into the overall metadata function. This theory argues for and presents an integrated mixed metadata approach rather than simply plugging in a specific technology in addition to existing standards-based metadata functions.

The theory posits that libraries should embrace the notion of enriching and filtering as an emerging paradigm and then decide what technology to use. Current metadata principles concern themselves with metadata that is created a priori by metadata experts. As the role of libraries as a physical space and book borrowers diminishes, libraries could re-envision themselves as knowledge building environments. In this re-envisioning process, it is argued, the Theory of Metadata Enriching and Filtering plays a significant role.

The theory succinctly delineates metadata content (enriching) from presentation (filtering) and about-ness from medium. The emergent theory of digital library metadata, although represented by two seemingly simple concepts, nonetheless elegantly abstracts the overarching concepts and principles developed in this thesis. The elegance of the theory lies in its ability to capture the complexity inherent in its multifaceted implications. By making the case for a mixed metadata approach, the theory also brings into focus the importance of re-conceptualising current metadata principles.

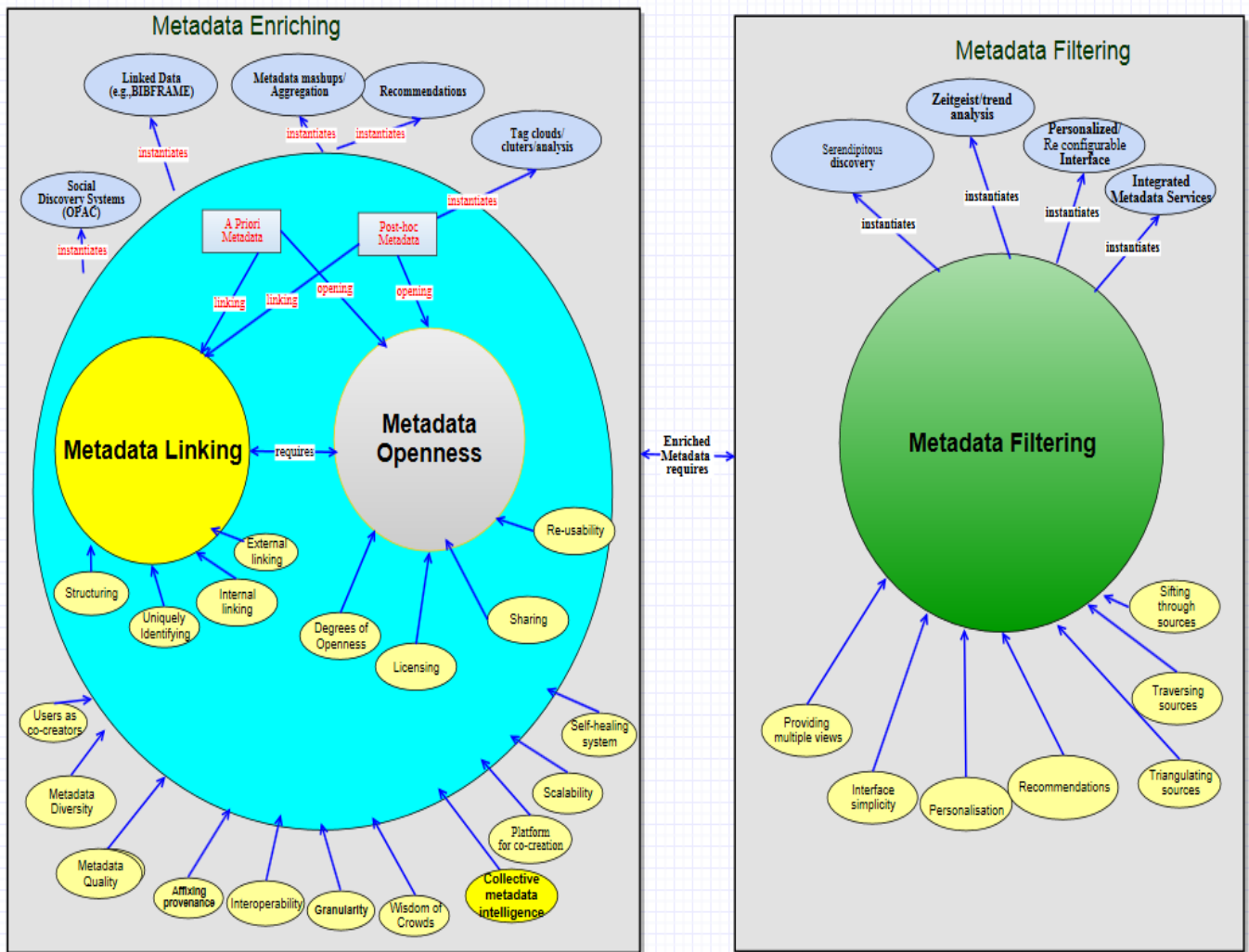


Figure 14.2: Relationship between the Principles of Enriching and Filtering

The principle of enriching requires the sub-principles of linking and opening (Figure 14.2) and each principle includes its own constituent concepts. Note that each of these concepts is discussed in Chapters Nine through Thirteen. By abstracting technologies, concepts, and principles, the Theory of Enriching and Filtering, it is argued, has a long lasting implication for digital library metadata. Enriching is a conceptual goal and it can only be supported through a mixed metadata paradigm, as the strategies that effect this paradigm require the consideration of the user as a proactive co-creator and the librarian as a metadata content architect.

Chapter Fifteen: Conclusions

15.1. Chapter Overview

The contributions of this thesis to knowledge are manifold. It has a theoretical contribution (15.2) as it develops novel concepts and principles, namely, enriching, linking, openness and filtering), which can underpin the creation and utilisation of metadata in libraries. By integrating these principles, an emergent Theory of Metadata Enriching and Filtering was developed (15.2.1). The theory suggests the emergence of a new metadata paradigm which calls for the re-consideration of librarian and user roles in libraries (15.2.2) and the re-conceptualisation of metadata quality (15.2.3). The theory also has practical considerations (15.3) with regard to the implications of these principles for standards-setting agencies (15.3.1), libraries (15.3.2), library systems developers (15.3.3) and library users (15.3.4).

15.2. Contributions to Theory

In this thesis, using a Constructivist Grounded Theory Method, four overarching metadata Core Categories, also called principles, emerged, namely metadata enriching (Chapter Nine), metadata linking (Chapter Ten), metadata openness (Chapter Eleven) and metadata filtering (Chapter Twelve). The integration of these Core Categories resulted in the emergence of a **Theory of Metadata Enriching and Filtering** (Chapter Fourteen).

The theory is considered an original contribution to the development, extension and refinement of contemporary metadata assumptions and principles. The theory provides a holistic framework demonstrating the interdependence between expert curated and socially-constructed metadata, wherein the former helps to structure the latter, whilst the latter provides diversity to the former. The theory elegantly presents several overarching complex concepts and principles subsumed under a mixed metadata approach.

15.2.1. An Emergent Theory of Metadata Enriching and Filtering

The Theory of Metadata Enriching and Filtering underpinned by four overarching metadata principles, namely, metadata enriching, linking, openness and filtering, is suggested. Whilst the concepts of linking, openness and filtering exist in existing literature, there lacks a holistic integration of these concepts. Thus, by providing a holistic theory and demonstrating the interdependence between these principles, this thesis develops and presents a novel metadata theory, which is considered to have wider implications in the way metadata is created, utilised and managed in libraries (see Figure 15.1).

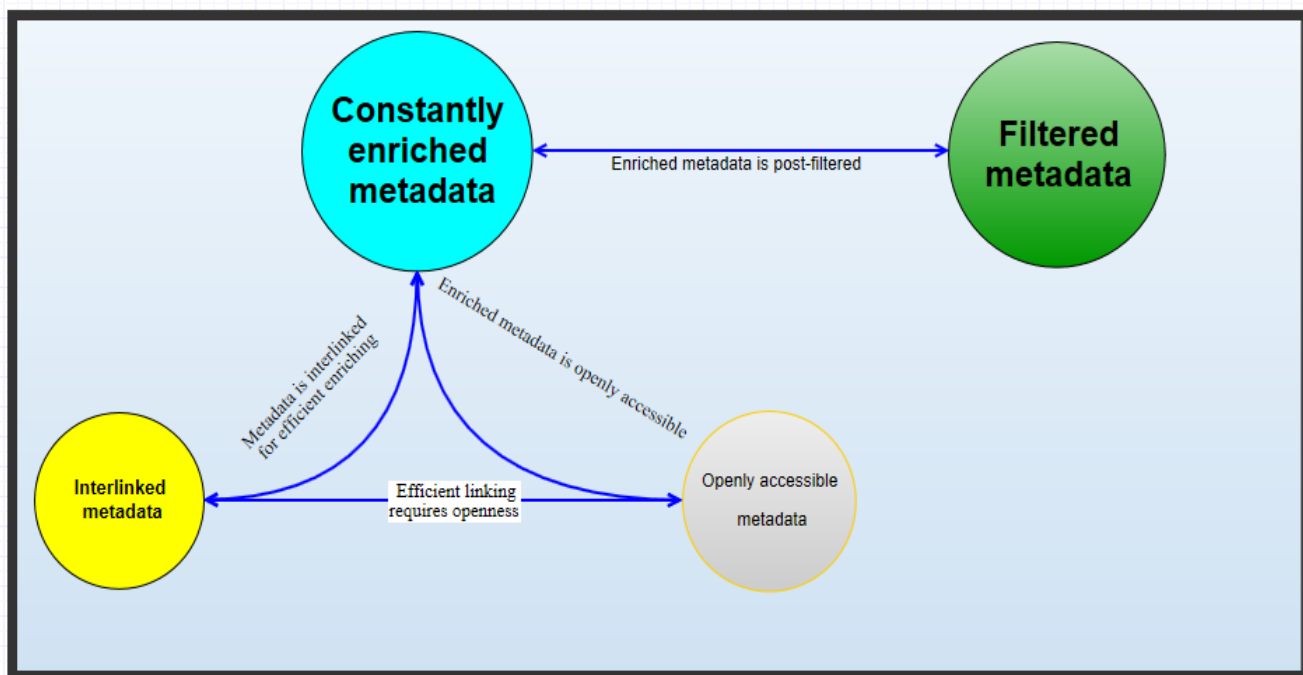


Figure 15.1: The Theory of Metadata Enriching and Filtering

The emergence of the Theory of Metadata Enriching and Filtering contributes to the development of the discipline of Library and Information Science in general and the field of digital libraries and metadata in particular, thus partly addressing the absence of theoretical foundations indicated in the metadata literature (Andersen & Skouvig, 2006; Day, 2010; Floridi, 2000; Lehmann, 2010) (see also 1.1; 6.11).

15.2.2. Re-Consideration of Librarian and User Roles

Central to the Theory of Metadata Enriching and Filtering is the consideration of users as pro-active metadata creators (8.4.2) rather than mere consumers (see Figure 15.2). Thus librarians become the experts at providing structure, granularity and interoperability to post-hoc metadata. In light of a mixed metadata approach, the consideration of socially-constructed (post-hoc) metadata for enriching standards-based (a priori) metadata assumes an interpretative ontological point of view, wherein multiple interpretations can be ascribed to an information object. The “social” in the phrase “socially-constructed metadata” implies the aggregated value, hence collective metadata intelligence (8.4.5), which can be garnered from the collective effort of users. The platform for metadata co-creation offers users the opportunity to add their own metadata describing information objects, in order to maximise the re-find-ability and discoverability of information objects at a later time. This is in sharp contrast with expert-curated metadata where it is dominated by terminologies which reflect only the librarian’s view point.

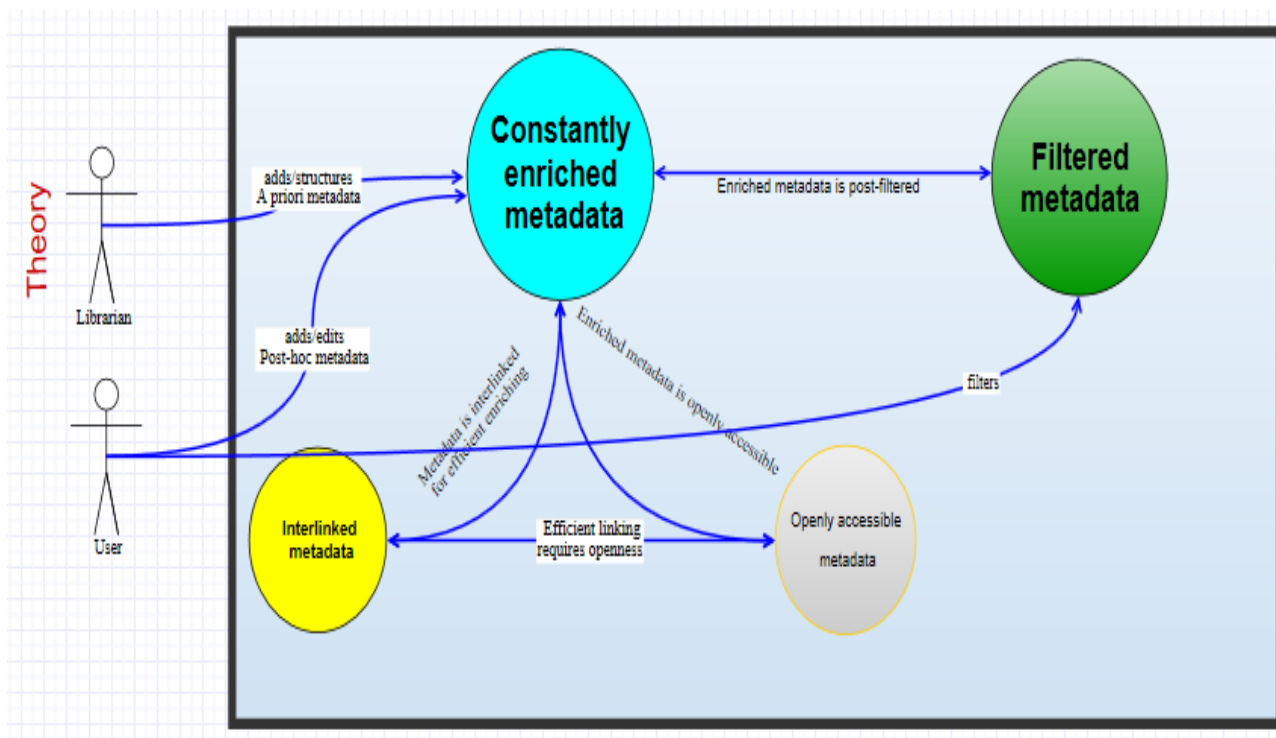


Figure 15.2: Changing Librarian and User Roles

15.2.3. Re-Conceptualisation of Metadata Quality

The Theory of Metadata Enriching and Filtering brings forth the importance of re-conceptualising the notion of metadata quality, stating that in a mixed metadata approach the issue of metadata quality (8.6) should be weighed in light of the relevance of metadata to support findability and discoverability. Conversely, it is metadata's usefulness for finding and discovering information objects, rather than on whether it is objectively accurate (truthful) or not, or on whether it is 'good enough', rather than on whether it is 'perfect metadata' that should be considered. In view of users as co-creators, it is considered essential that users are involved not only in the creation of metadata but also in its maintenance and management, thereby creating a self-monitoring and self-healing system (8.4.5.7).

15.3. Contributions to Practice

The Theory of Metadata Enriching and Filtering has profound practical implications for standards-setting agencies (15.3.1), libraries (15.3.2), library management system developers (15.3.3) and library users (15.3.4).

15.3.1. Implications for Standards-Setting Agencies

The Theory of Metadata Enriching and Filtering can help inform standards-settings authorities and can be of benefit to initiatives aimed at a re-conceptualisation of contemporary metadata principles. The principle of metadata linking is bound to entail that multiple agencies will need to work in close cooperation, for example when assigning unique identifiers to metadata (persistent identifiers). The theory also requires standards agencies to

design and develop adapting and evolving standards that can easily accommodate changes, such as the incorporation of users' terminologies and new metadata fields.

15.3.2. Implications for Libraries

The theory is expected to have practical implications on libraries in the following areas (see Figure 15.3):

- From metadata simplicity to metadata enriching;
- From human-readable and non-linked metadata to granularly structured, uniquely identified, machine process-able and interlinked metadata;
- From metadata that is locked to metadata openness;
- From single (global) OPAC interfaces to re-configurable interfaces (filtering).

15.3.2.1. From Metadata Simplicity to Metadata Enriching

Libraries should consider socially-constructed metadata approaches as a novel opportunity to re-orient their metadata functions to meet users' needs. In this regard, libraries should be cognisant of the fact that metadata contributors in a socially-constructed metadata approach can possibly be people of expertise in their subject domains and, thereby be valued contributors, for a number of motivations including prestige, altruism or just concern for one's field (8.5.5). In this regard, since librarians are not always experts in subject domains, involving users enables libraries to solicit valuable metadata with regard to the about-ness of information objects. Since post-hoc metadata is created by users themselves, the likelihood to match their information search keywords is much higher than if standardised and formal taxonomies are used (8.5.4 & 8.5.6).

15.3.2.2. From Non-Linked to Interlinked Metadata

The principle of metadata linking, which emerged in this thesis as an important consideration for libraries, would address some of the major limitations of current standards-based metadata approaches. The data analysis revealed that, current library metadata standards are not designed for linking. Thus a lot of future work is needed. The principle of linking can obviate these challenges by providing a granular, machine process-able, re-usable and link-able structure (see 10.4). These benefits can be used in a mixed metadata approach. Thus, the metadata diversity (8.4.3) that resulted from the adoption of socially-constructed metadata approaches can be effectively structured, affixed with provenance and inter-linked (see Chapter Ten). In this regard, current metadata standards and models such as RDA and FRBR should be re-conceptualised, so that it can be retrofitted to the principle of linking.

15.3.2.3. From Metadata that is Locked (Silos) to Metadata Openness

Embracing the principle of metadata openness helps libraries to look beyond their institutional silos, further benefitting themselves through metadata sharing, re-use, mixing and matching and integration. Metadata openness significantly reduces the cost of metadata

creation by avoiding duplication of effort. However, issues of licensing need to be addressed. Metadata created for a specific purpose could thus be re-purposed and re-used to achieve a completely different purpose in a manner that may not have been foreseen by the original creator. In addition, as the data analyses revealed, metadata openness furthers institutional transparency and accountability. Finally, metadata openness, rather than guarding metadata in institutional silos, increases the returns on investment expended on metadata creation. Hence, libraries should proactively coordinate and implement open data policies and strategies.

15.3.2.4. From a Single (Global) OPAC Interface to Re-configurable, Multi-layered Interfaces

As discussed in Chapter Twelve and Thirteen, both LIS professionals and library users indicated the importance of re-orienting library search and discovery interfaces towards personalised, customisable and re-configurable displays (interfaces) rather than the traditional single global interface.

15.3.3. Implications in Library Management Systems (LMS) Software

In relation to discovery interfaces such as the OPAC, providing multiple metadata views (12.2), interface simplicity (12.3), personalisation (12.4), recommendations (12.5; 12.6), integrated access to information services (12.7), serendipitous access (12.8), convenience of access (13.4.1), the ability to traverse (13.3.1), sift (13.3.2) and triangulate (13.3.3), are all important considerations in the design of user-friendly Library Management Systems (LMS) software.

15.3.4. Implications for Users

The Theory of Metadata Enriching and Filtering has indispensable benefit for users. First of all, socially-constructed metadata platforms (8.4.1) enable users to be pro-active metadata creators (8.4.2), which in turn provides metadata diversity (8.4.3), reflecting thus the various terminologies and interpretations of users. The role of users therefore shifts from being passive consumers to be active partners in the creation and management of metadata. The more a metadata system reflects the language (vocabularies/terminologies) of users, the better it is likely to support the findability and discoverability of information objects in libraries.

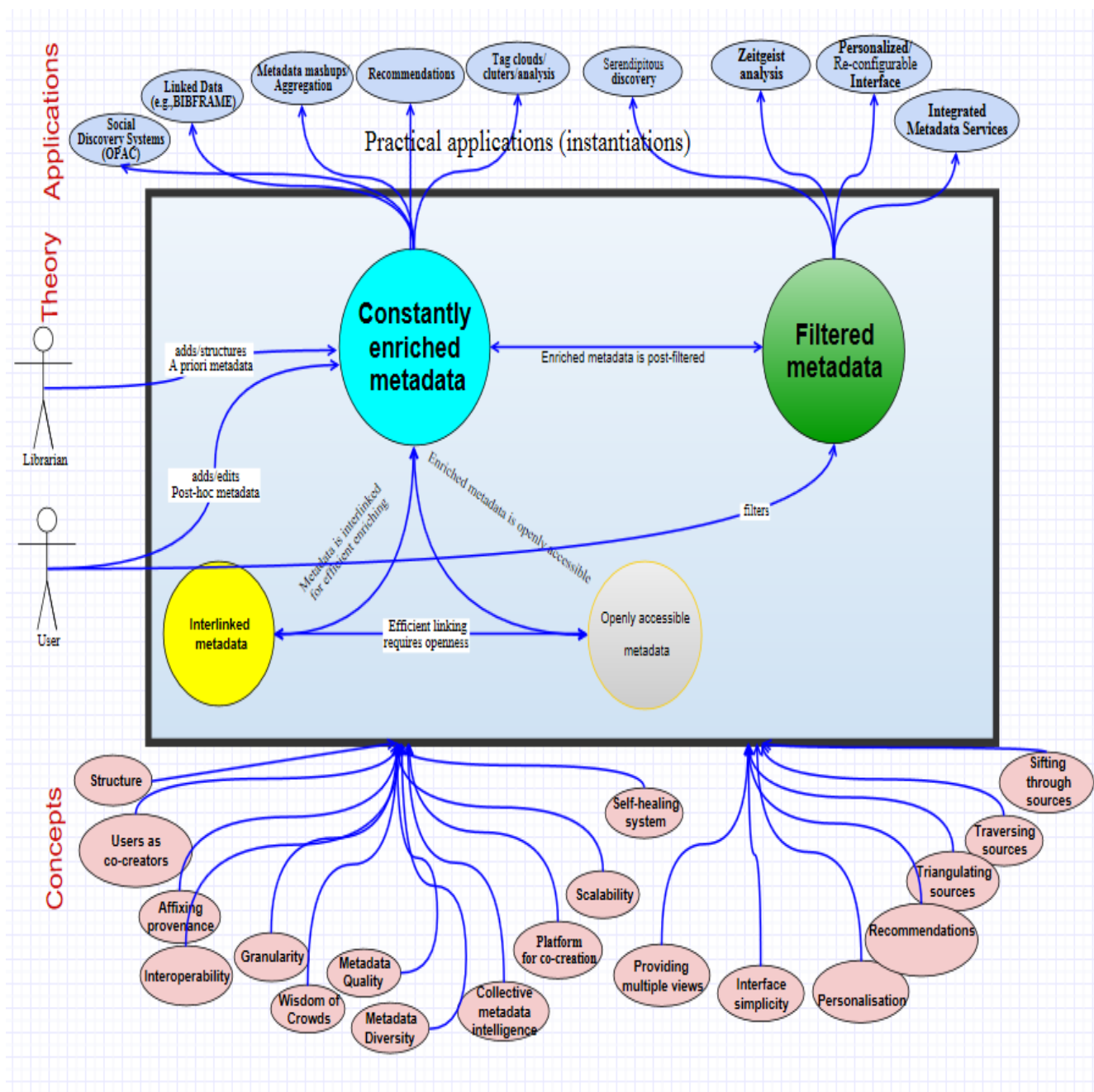


Figure 15.3: Practical Implications of the Theory of Metadata Enriching and Filtering

15.4. Methodological Contribution

The social constructivist approach employed in this study is a relatively novel methodology for investigating and understanding emerging metadata issues, identifying relevant concepts, principles and developing a digital library metadata theory. Adopting a social constructivist approach, it is argued, enables one to better reflect the diversity in needs, capabilities and experiences that exists among stakeholders and library users. In light of the novelty of the adoption of the Web 2.0 paradigm in metadata functions, the availability of existing theories and principles is scant, thus an inductive rather than deductive research method was considered appropriate. This thesis thus used a Constructivist Grounded Theory Methodology, which was found appropriate for exploring emerging metadata approaches.

The diversity of the experiences, authority and expertise of participants, along with the in-depth interviewing approach that has been employed, augmented by the rigour with which Constructivist Grounded Theory analytic procedures were followed, and, finally, the thoroughness of coding and memo writing practices, leads one to conclude that the required depth and rigour have been accomplished in this research. This method offered depth and rigour to identify and develop concepts and principles. Thus, the consistent adoption of a social constructivist (interpretive) philosophical view point and a Constructivist Grounded Theory Method is considered an important methodological contribution to investigate emerging research topics in LIS.

15.5. Limitations of this Research

Whilst it has been clear that the Constructivist Grounded Theory Methodology has worked well for identifying emergent topics pertaining to the combined use of socially-constructed and standards-based metadata approaches, it should be admitted that this is a very time consuming process, involving several iterative instances of collecting data using interviews and analysis using coding and theoretical saturation procedures. As it is an inductive process, identification of potential research participants had to be derived from preceding data analyses, thus requiring additional time for reflection and evaluation through memo writing before additional data could be collected.

In addition, even though the derived substantive theory can be used to explain and inform the creation and utilisation of the general category of descriptive metadata, the contexts chosen for this thesis are academic libraries and books. Most of the interviewees are also affiliated to university libraries. Thus it did not consider other types of metadata, such as administrative, preservation or structural metadata (as in Figure 1.1), which is a subject for future research (15.6). The focus of this thesis is on digital libraries and digital books. However, the theory developed in this thesis would have implications for domains other than libraries and other genres of information objects. It would, for example, be essential to contextualise this theory to open research data. Such expansion would reinvigorate the role of libraries in supporting the dissemination and effective utilisation of academic research.

15.6. Future Research

Charmaz (2006) and Strauss and Corbin (1998) advise that grounded theory research should be validated by future research. Thus, it is important that future research takes the Theory of Enriching and Filtering and develop specific models and experiments with incorporating them into the development of library metadata systems and interfaces (OPAC/discovery systems). Once such models are developed, it would also be pertinent to investigate and evaluate how such systems benefit users and add value to the practical findability and discoverability of information objects in libraries. It is also important that future research investigates the technological ways of implementing the Theory of Metadata Enriching and Filtering.

However, the grounded theory method, through an inductive analysis, resulted in the Theory of Metadata Enriching and Filtering. Even though the grounded theory methodology adopted

in this research does not necessitate validation by the same researcher with in the same time frame per se, this research could benefit from a post-hoc survey of librarians as to the viability of implementing the emergent principles in an actual library setting. Thus, in future research, the theory can be tested using a deductive methodology before models and systems are developed.

Whilst the scope of this thesis is restricted to descriptive metadata in the context of academic libraries, the theory emerged from this thesis can have wider implications to the area of digital library metadata, such as administrative, structural and preservation metadata. In addition, if extended and confirmed through future research, it can also have applications to areas outside the library domain, such as government data and commercial data. In other words, the Theory of Enriching and Filtering can have implications for other genres of information objects and other types of data other than library metadata. Future research would therefore help to contextualise the results of this thesis to other areas.

15.7. Concluding Remarks

The journey towards the discovery of the grounded Theory of Metadata Enriching and Filtering has been a challenging but rewarding endeavour. Iterative data collection, analysis and inductive build-up of concepts and principles from empirical data guided the research in paths that the researcher had not fully anticipated from the start. However, the emergence of a novel digital library metadata theory is a rewarding one.

Enrich then **filter**. This is the emergent theory. A seemingly simple yet elegant representation of the complexities embedded in the overarching concepts, principles and technologies that are essential for a new metadata approach. The theory reveals the importance of delineating metadata content from its presentation. It reveals that metadata enriching is a continual and non-deterministic process. It is argued that the theory explains, informs and guides the development, extension and refinement of contemporary metadata assumptions and principles.

The theory posits the acceptance and inclusion of diverse perspectives and interpretations on matters of metadata description. Whilst the theory does not object to the existence of verifiable entities (realities), it however espouses that our understanding of these objects does not always coalesce into a single interpretation. This is especially true when a multitude of users are involved as metadata co-creators. This research demonstrated the credibility of adopting a social constructivist philosophical approach as well as a Constructivist Grounded Theory Method for investigating emergent topics. It is also argued that the findings of this thesis resonate well with the data; are credible, ‘original’ and useful interpretations, but the reader should bear in mind, in accordance with the social constructivist philosophical approach, that the theory recommended in this research is one of many possible interpretations. It is thus suggested that future research should revisit, refine, and extend it, in short the theory itself should be continually enriched.

References

- Adams, C., & Ramos, I. (2010). *The Past, Present and Future of Social Networking and Outsourcing: Impact on Theory and Practice*. Paper presented at the UK Academy for Information Systems Conference, Oriel College, University of Oxford.
- Alemneh, D. G. (2009). *An Examination of the Adoption of Preservation Metadata in Cultural Heritage Institutions: An Exploratory Study Using Diffusion of Innovations Theory* (PhD Thesis), University of North Texas. Retrieved from <http://digital.library.unt.edu/ark:/67531/metadc9937/?q=gelaw>
- Alemu, G., Stevens, B., & Ross, P. (2012a). Semantic Metadata Interoperability in Digital Libraries: A Constructivist Grounded Theory Approach. *Bulletin of IEEE Technical Committee on Digital Libraries*, 8(1).
- Alemu, G., Stevens, B., & Ross, P. (2012b). Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: A social constructivist approach. *New Library World*, 113(1/2), 38-54.
- Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012). Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. *New Library World*, 113(11/22), 549-570.
- Alexander, B. (2006). Web 2.0: A New Wave of Innovation for Teaching and Learning? *EDUCAUSE Review*, 42(2), 32-44.
- Allan, G. (2003). A critique of using grounded theory as a research method. *Electronic Journal of Business Research Methods*, 2(1), 1-10.
- Allan, G. (2007). *The Use of the Grounded Theory Methodology in Investigating Practitioners Integration of COTS Components in Information Systems*. Paper presented at the Twenty Eighth International Conference on Information Systems, Montréal, Canada.
- Allemnag, D., & Hendler, J. (2008). *Semantic Web for the working ontologist: effective modeling in RDFS and OWL*. Amsterdam: Morgan Kaufmann.
- Andersen, J., & Skouvig, L. (2006). Knowledge organization: a sociohistorical analysis and critique. *Library Quarterly*, 76(3), 300-322.
- Anderson, C. (2006). *The long tail: how endless choice is creating unlimited demand*. London: Random House Business Books.
- Anderson, C. (2009a). *Free : the future of a radical price*. New York: Hyperion.
- Anderson, C. (2009b). *The longer long tail : how endless choice is creating unlimited demand* (Expanded & updated ed.). London: Random House Business Books.
- Anderson, D., Delve, J., Pinchbeck, D., & Alemu, G. A. (2009). Preservation Metadata Standards for Emulation Access Platforms. Portsmouth, UK: KEEP Project, European Commission.
- Anderson, P. (2007). What is Web 2.0? Ideas, technologies and implications for education *JISC Technology & Standards Watch Report*: JISC
- Arms, W., Hillmann, D., Lagoze, C., Krafft, D., Marisa, R., Saylor, J., Van de Sompel, H. (2002). A Spectrum of interoperability: the site for science prototype for the NSDL. *D-Lib Magazine*, 8.
- Atherton, A., & Elsmore, P. (2007). Structuring qualitative enquiry in management and organization research: A dialogue on the merits of using software for qualitative data analysis. *Qualitative Research in Organizations and Management*, 2(1), 62-77.
- Bailer, W. (2007). *Interoperability of Multimedia Metadata: from Digital Cinema to Cultural Heritage*. Paper presented at the CHORUS Workshop on Metadata in Audio-Visual/Multimedia Productions and Archiving, Munich, Germany. http://www.ist-chorus.org/_events_RTf/documents/WSChorus_Bailer.pdf
- Baker, T. (1998). Languages for Dublin Core. *D-Lib Magazine*, 4(12).
- Baker, T. (2000). A Grammar of Dublin Core. *D-Lib Magazine*, 6(10).
- Barbosa, D. (2008). *Taxonomy Folksonomy Cookbook* Retrieved from http://solutions.dowjones.com/cookbook/ebook_sla2008/cookbookebook.pdf
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5), 407-424. doi: 10.1108/eb024320

- Bearman, D., Miller, E., Rust, G., Trant, J., & Weibel, S. (1999). A Common Model to Support Interoperable Metadata: Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities. *D-Lib Magazine*, 5(1).
- Berners-Lee, T. (1997). Metadata Architecture. <http://www.w3.org/DesignIssues/Metadata.html>
- Berners-Lee, T. (1998). What the semantic web can represent. <http://www.w3.org/DesignIssues/RDFnot.html>
- Berners-Lee, T. (2009). Linked Data. <http://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *The Scientific American*, 284(5), 34-43
- Blair, A. (2010). *Too much to know: managing scholarly information before the modern age*. New Haven, Conn.: Yale University Press.
- Blanchi, C., & Petrone, J. (2001). Distributed Interoperable Metadata Registry. *D-Lib Magazine*, 7(12).
- Borgman, C. L. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science*, 47(7), 493-503.
- Bowker, G., & Star, S. L. (1999). *Sorting things out: classification and its consequences*. Cambridge, Mass: MIT Press.
- Britannica Inc. (2006). Fatally Flawed: Refuting the recent study on encyclopedic accuracy by the journal Nature. http://corporate.britannica.com/britannica_nature_response.pdf
- Brown, J. S., & Duguid, P. (2000). *The social life of information*. Boston, Mass: Harvard Business School.
- Bruce, T., & Hillmann, D. (2004). The Continuum of Metadata Quality: Defining, Expressing, and Exploiting. In D. I. Hillmann & E. L. Westbrooks (Eds.), *Metadata in Practice* (pp. 238-256). Chicago: American Library Association.
- Bryant, A., & Charmaz, K. (2007a). Grounded theory in historical perspective: an epistemological approach. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 31-57). Los Angeles: SAGE Publications.
- Bryant, A., & Charmaz, K. (2007b). Introduction grounded theory research: methods and practises. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 1-28). Los Angeles: SAGE Publications.
- Buchanan, G., & McKay, D. (2011). *In the bookshop: examining popular search strategies*. Paper presented at the JCDL-2011, 11th annual international ACM/IEEE joint conference on Digital libraries, Ottawa, Canada.
- Buckland, M. (1999). *Vocabulary as a Central Concept in Library and Information Science*. Paper presented at the Third International Conference on Conceptions of Library and Information Science Dubrovnik, Croatia.
- Cabinet Office. (2004). e-Government Interoperability Framework. http://edina.ac.uk/projects/interoperability/e-gif-v6-0_.pdf
- Calhoun, K. (2006). The Changing Nature of the Catalog and its Integration with Other Discovery Tools: final report prepared for the Library of Congress (pp. 52). Ithaca, N.Y: Cornell University Library.
- Caplan, P. (2000). Oh what a tangled web we weave: Opportunities and challenges for standards development in the digital library arena. *First Monday*, 5(6).
- Caplan, P. (2003). *Metadata fundamentals for all librarians*. Chicago: American Library Association.
- Caplan, P. (2009). Understanding PREMIS. Washington D.C: The Library of Congress.
- Casey, M. E., & Savastinuk, L. C. (2006). Web 2.0: service for the next-generation library. *Library Journal*. <http://www.libraryjournal.com/article/CA6365200.html>
- Casey, M. E., & Savastinuk, L. C. (2007). *Library 2. 0: a guide to participatory library service*. Medford, N.J: Information Today, Inc.
- CC:DA. (2000). Committee on Cataloging: Description & Access Task Force on Metadata: Final Report: American Library Association.
- Chan, L. M., & Zeng, M. L. (2006). Metadata Interoperability and Standardization – A Study of Methodology Part I :Achieving Interoperability at the Schema Level. *D-Lib Magazine*, 12(6).

- Charmaz, K. (2006). *Constructing grounded theory: a practical guide through qualitative analysis*. London: SAGE Publications.
- Choy, F. C. (2011). From library stacks to library-in-a-pocket: will users be around? *Library Management*, 32(1/2), 62-72.
- Christakis, N. A., & Fowler, J. H. (2011). *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. London: Harper Press.
- Cobden, M., Black, J., Gibbins, N., Carr, L., & Shadbolt, N. (2011). *A Research Agenda for Linked Closed Data*. Paper presented at the Second International Workshop on Consuming Linked Data, Bonn, Germany. <http://eprints.ecs.soton.ac.uk/22711/3/position.pdf>
- Coyle, K. (2009). Metadata Mix and Match. *Information Standards Quarterly*, 21(1), 9-11.
- Coyle, K. (2010). Library Data in a Modern Context. *Library Technology Reports*, 46(1), 5-13.
- Coyle, K., & Hillmann, D. (2007). Resource Description and Access (RDA): Cataloging Rules for the 20th Century. *D-Lib Magazine*, 13(1/2).
- Crotty, M. (1998). *The Foundations of Social Research: Meaning and Perspective in the Research Process*. London: SAGE Publications.
- Dappert, A., & Enders, M. (2008). Using METS, PREMIS and MODS for Archiving eJournals. *D-Lib Magazine*, 14(9/10).
- Day, M. (2000). Resource discovery, interoperability and digital preservation: some aspects of current metadata research and development. *VINE*, 36(117), 35-48.
- Day, M. (2001). Metadata in a nutshell. *Information Europe*, 6(2), 11.
- Day, M. (2003a). *Integrating metadata schema registries with digital preservation systems to support interoperability: a proposal*. Paper presented at the International Conference on Dublin Core and Metadata Applications, Seattle.
- Day, M. (2003b). *Preservation metadata initiatives: practicality, sustainability, and interoperability*. Paper presented at the ERPANET Training Seminar on Metadata in Digital Preservation, Marburg, Germany. <http://www.ukoln.ac.uk/preservation/publications/erpanet-marburg/day-paper.pdf>
- Day, M. (2005). DCC Digital Curation Manual Instalment on Metadata.
- Day, R. (2010). The Self-Imposed Limits of Library and Information Science: Remarks On the Discipline, On the Profession, On the University, and On the State of "Information" in the U.S. at Large Today. *InterActions: UCLA Journal of Education and Information Studies*, 6(2).
- Decker, S., Melnik, S., Van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., . . . Horrocks, I. (2000). THE SEMANTIC WEB: The Roles of XML and RDF. *IEEE Internet Computing*, 15(3), 63-74.
- Dekkers, M. (2002). Issues in cross-standard interoperability. www.cores-eu.net/interoperability/d31.pdf
- Dekkers, M., & Weibel, S. (2002). Dublin Core Metadata Initiative Progress Report and Workplan for 2002. *D-Lib Magazine*, 8(2).
- Dempsey, L. (2004). *Interoperability: the value of recombinant potential* Paper presented at the ARLIS 2004: the 32nd Annual Conference of the Art Libraries Society of North America New York City <http://www.oclc.org/research/presentations/dempsey/arlis04.ppt>
- Denton, W. (2007). FRBR and the History of Cataloging. In A. G. Taylor (Ed.), *Understanding FRBR: What It Is and How It Will Affect Our Retrieval* (pp. 35-57). Westport, Connecticut: Libraries Unlimited.
- Doctorow, C. (2001). Metacrap: Putting the torch to seven straw-men of the meta-utopia. <http://www.well.com/~doctorow/metacrap.htm>
- Duffy, T. M., & Jonassen, D. H. (Eds.). (1992). *Constructivism and the technology of instruction: a conversation*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.
- Dunn, L. (2011). *An investigation of the factors affecting the lifecycle costs of COTS-based systems*. (PhD), University of Portsmouth Retrieved from http://eprints.port.ac.uk/4135/1/Laurence_Dunn_Theses_Document_FINAL_v11_06JUN11.pdf

- Dunsire, G. (2008). *The Semantic Web and expert metadata: pull apart then bring together* Paper presented at the Archives, Libraries, Museums (AKM12), Poreč, Croatia.
<http://eprints.rclis.org/bitstream/10760/7443/1/akm2008semanticweb.pdf>
- Dunsire, G. (2009). *UNIMARC, RDA and the Semantic Web* Paper presented at the World Library and Information Congress: 75th IFLA General Conference and Council Milan, Italy.
<http://www.ifla.org/files/hq/papers/ifla75/135-dunsire-en.pdf>
- Dunsire, G. (2009). *UNIMARC, RDA and the Semantic Web* Paper presented at the World Library and Information Congress: 75th IFLA General Conference and Council Milan, Italy.
<http://www.ifla.org/files/hq/papers/ifla75/135-dunsire-en.pdf>
- Dunsire, G. (2012). Representing the FR Family in the Semantic Web. *Cataloging & Classification Quarterly*, 1-18.
- Dunsire, G., & Willer, M. (2011). Standard library metadata models and structures for the Semantic Web. *Library Hi Tech News*, 28(3), 1-12.
- Duval, E., Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata Principles and Practicalities. *D-Lib Magazine*, 8(4).
- EC. (2004). European Interoperability Framework for Pan-European e-Government Services. Retrieved 09 September 2009, from <http://ec.europa.eu/idabc/servlets/Doc?id=19528>
- EC. (2010). European Interoperability Framework for Pan-European e-Government Services. Retrieved 24 March 2011, from <http://ec.europa.eu/idabc/servlets/Docd552.pdf?id=19529>
- Evans, W. (2009). *Building library 3.0: issues in creating a culture of participation*. Oxford: Chandos.
- Farkas, M. G. (2007). *Social software in libraries: building collaboration, communication, and community online*. Medford, N.J: Information Today, Inc.
- Fifarek, A. (2007). The Birth of Catalog 2.0: Innovative Interfaces' Encore Discovery Platform. *Library Hi-Tech News* (5), 13-15. doi: 10.1108/07419050710780353
- Floridi, L. (2000). On defining library and information science as applied philosophy of information. *Social Epistemology*, 16(1), 37-49.
- Floridi, L. (2009). Web 2.0 vs. the Semantic Web: A Philosophical Assessment. *Episteme*, 6(1), 25-37.
- Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science*, 30(4), 378-391.
- Gartner, R. (2008). Metadata for digital libraries: state of the art and future directions. *JISC Technology and Standards Watch Reports: Bristol*.
http://www.jisc.ac.uk/media/documents/techwatch/tsw_0801pdf.pdf
- Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438.
<http://www.nature.com/nature/journal/v438/n7070/pdf/438900a.pdf>
- Gill, T., & Miller, P. (2002). Re-inventing the Wheel? Standards, Interoperability and Digital Cultural Content. *D-Lib Magazine*, 8(1).
- Glaser, B. (2001). *The Grounded Theory Perspective I: Conceptualization Contrasted with Description*. Mill Valley, CA: Sociology Press.
- Glaser, B. (2002). Constructivist Grounded Theory? *Forum: Qualitative Social Research*, 3(3).
- Glaser, B. G. (1978). *Theoretical sensitivity : advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory : strategies for qualitative research*. London: Weidenfeld and Nicolson.
- Greenberg, J. (2005). Understanding Metadata and Metadata Schemes. *Cataloging & Classification Quarterly*, 40(3-4), 17-36.
- Greenberg, J., Sutton, S., & Campbell, D. G. (2005). Metadata: A Fundamental Component of the Semantic Web. *Bulletin of the American Society for Information Science and Technology*, 29(4), 16-18.
- Gruber, T. (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies*, 43, 907-928.
- Gruber, T. (2007). Ontology of folsonomy: a mash-up of apples and oranges. *International Journal on Semantic Web & Information Systems*, 3(2).
- Guba, G., & Lincoln, S. (1989). *Fourth Generation Evaluation*. London: SAGE Publications.

- Guba, G., & Lincoln, S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: SAGE.
- Guenther, R. (2005). Preservation Metadata: Implementation Strategies (PREMIS).
- Guenther, R., & McCallum, S. (2003). New metadata standards for digital resources: MODS and METS. *Bulletin of the American Society for Information Science and Technology*, 29(2).
- Guenther, R. S. (2008). Battle of the Buzzwords: Flexibility vs. Interoperability When Implementing PREMIS in METS. *D-Lib Magazine*, 14(7/8).
- Guy, M. (2006). Folksonomies: Tidying up Tags? . *D-Lib Magazine*, 12(1).
- Haslhofer, B., & Klas, W. (2010). A Survey of Techniques for Achieving Metadata Interoperability. *ACM*, 42(2).
- Haynes, D. (2004). *Metadata for information management and retrieval*. London: Facet.
- Hedden, H. (2010). *The accidental taxonomist*. Medford, N.J: Information Today.
- Heery, R., & Patel, M. (2000). Application profiles: mixing and matching metadata schemas. *Ariadne* (25). <http://www.ariadne.ac.uk/issue25/app-profiles/>
- Hjorland, B. (2000). Library and information science: practice, theory, and philosophical basis. *Information Processing and Management* 36, 501-531.
- Hoffman, G. L. (2009). Meeting Users' Needs in Cataloging: What is the Right Thing to Do? *Cataloging & Classification Quarterly*, 47(7), 631-641.
- Hurley, B. J., Price-Wilkin, J., Proffitt, M., & Besser, H. (1999). The Making of America II Testbed Project: a digital library service model. Washington, DC: The Digital Library Federation (DLF), Council on Library and Information Resources (CLIR).
- IFLA. (2009). Functional Requirements for Bibliographic Records: Final Report: International Federation of Library Associations and Institutions.
- Keen, A. (2007). *The Cult of the Amateur: How Today's Internet Is Killing Our Culture and Assaulting Our Economy*: Nicholas Brealey Publishing.
- Kroski, E. (2005). The Hive Mind: Folksonomies and User-Based Tagging. In E. Kroski (Ed.), *Infotangle*.
- Kroski, E. (2008). *Web 2.0 for librarians and information professionals*. New York: Neal-Schuman.
- Kuhn, T. S. (1962). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press.
- Kukla, A. (2000). *Social constructivism and the philosophy of science*. London: Routledge.
- Lagoze, C. (2001a). *Accommodating Simplicity and Complexity in Metadata: Lessons from the Dublin Core Experience* Paper presented at the Seminar on Metadata, Amsterdam.
- Lagoze, C. (2001b). Keeping Dublin Core Simple: Cross-Domain Discovery or Resource Description? *D-Lib Magazine*, 7(1).
- Lagoze, C. (2010). *Lost Identity: the Assimilation of Digital Libraries into the Web*. (PhD), Cornell University, School of Information Science. Retrieved from <http://www.cs.cornell.edu/lagoze/dissertation/CarlLagoze.pdf>
- Lagoze, C., & Patzke, K. (2011). *A research agenda for data curation cyberinfrastructure*. Paper presented at the The 11th annual international ACM/IEEE joint conference on Digital libraries - JCDL '11, Ottawa, Canada.
- Lavoie, B., & Gartner, R. (2005). Preservation Metadata *DPC Technology Watch Series Report 05-01*: Online Computer Library Center Inc.(OCLC).
- Lavoie, B. F. (2004). Implementing Metadata in Digital Preservation Systems: The PREMIS Activity. *D-Lib Magazine*, 10(4).
- Lavoie, B. F. (2008). PREMIS With a Fresh Coat of Paint: Highlights from the Revision of the PREMIS Data Dictionary for Preservation Metadata. *D-Lib Magazine*, 14(5/6).
- Lawrence, G. W., Kehoe, W. R., Rieger, O. Y., Walters, W. H., & Kenney, A. R. (2000). *Risk Management of Digital Information: A File Format Investigation*. Washington, D.C.: Council on Library and Information Resources.
- Lehmann, H. (2010). *Grounded Theory and Information Systems: Are We Missing the Point?* Paper presented at the The 43rd Hawaii International Conference on System Sciences, Koloa, Kauai, Hawaii.

- Lu, C., Park, J., & Hu, X. (2010). (2010). User tags versus expert-assigned subject terms: A comparison of Librarything tags and Library of Congress subject headings. *Journal of Information Science*, 36(6), 763–779.
- Lubetzky, S. (1953). Development of Cataloging Rules *Library Trends*, 2(2), 179-186.
- Lubetzky, S. (1955). Development of Cataloging Rules. *Library Trend*, 2(2), 179-186.
- Maness, J. M. (2006). Library 2.0 Theory: Web 2.0 and Its Implications for Libraries. *Webology*, 3(2).
- Mansourian, Y. (2006). Adoption of grounded theory in LIS research. *New Library World*, 107(1228/1229), 386-402.
- Marcum, D. (2011). *A Bibliographic Framework for the Digital Age*. Washington, D.C.: Library of Congress.
- Marcum, D. B. (2005). *The Future of Cataloging* Paper presented at the Ebsco Leadership Seminar, Boston, Massachusetts <http://www.loc.gov/library/reports/CatalogingSpeech.pdf>
- Mathes, A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Merholz, P. (2004). Metadata for the Masses. *Adaptive Path*.
<http://www.adaptivepath.com/publications/essays/archives/000361.php>
- Miller, P. (2000). Interoperability What is it and Why should I want it? *Ariadne*, (24).
<http://www.ariadne.ac.uk/issue24/interoperability/intro.html>
- Miller, P. (2005). Web 2.0: Building the New Library *Ariadne*, (45).
<http://www.ariadne.ac.uk/issue45/miller/#1>
- Mills, J., Bonner, A., & Francis, K. (2006). Adopting a constructivist approach to grounded theory: Implications for research design. *International Journal of Nursing Practice*, 12, 8-13.
- Morville, P. (2005). *Ambient findability*. Sebastopol, CA: O'Reilly.
- Nagamori, M., & Sugimoto, S. (2006). A Metadata Schema Registry as a Tool to Enhance Metadata Interoperability. *TCDL Bulletin*, 3(1).
- Nguyen, L. C., Partridge, H., & Edwards, S. L. (2012). *Understanding the participatory library through a grounded theory study*. Paper presented at the The American Society for Information Science and Technology, Baltimore, MD.
- Nielsen, M. (2012). *Reinventing Discovery: The New Era of Networked Science*. Princeton, New Jersey: Princeton University Press.
- Nilsson, M. (2010). *From interoperability to harmonization in metadata standardization: designing an evolvable framework for metadata harmonization*. (PhD PhD), School of Computer Science and Communication, KTH, Stockholm. Retrieved from
<http://kmr.nada.kth.se/papers/SemanticWeb/FromInteropToHarm-MikaelsThesis.pdf>
- Nilsson, M., Johnston, P., Naeve, A., & Powell, A. (2006). *Towards an Interoperability Framework for Metadata Standards*. Paper presented at the International Conference on Dublin Core and Metadata Applications, Colima, Mexico.
- NISO. (2004). Understanding Metadata.
<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>
- NISO. (2007). *A Framework of Guidance for Building Good Digital Collections NISO Recommended Practices*. Baltimore, MD, USA: National Information Standards Organization (NISO).
- NLA. (1999). *Preservation Metadata for Digital Collections*.
- O'Reilly, T. (2005). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. <http://oreilly.com/web2/archive/what-is-web-20.html>
- O'Reilly, T., & Battelle, J. (2009). *Web Squared: Web 2.0 Five Years On*. Paper presented at the Web 2.0 Summit, San Francisco, CA.
http://assets.en.oreilly.com/1/event/28/web2009_websquared-whitepaper.pdf
- OCLC/RLG. (2008). PREMIS Data Dictionary for Preservation Metadata version 2.0.
- Ouksel, A. M., & Sheth, A. (1999). Semantic Interoperability in Global Information Systems: A brief introduction to the research area and the special section. *ACM SIGMOD Record*, 28(1), 5-12.
- Park, J.-R., & Tosaka, Y. (2010). Metadata Creation Practices in Digital Repositories and Collections: Schemata, Selection Criteria, and Interoperability. *Information Technology & Libraries* 29(3), 104-116.

- Peterson, E. (2006). Beneath the Metadata: Some Philosophical Problems with Folksonomy. *D-Lib Magazine*, 12(11).
- Pink, D. H. (2005, December 11). Folksonomy, *New York Times*. Retrieved from <http://www.nytimes.com/2005/12/11/magazine/1ideas1-21.html>
- Pressley, L. (2005). *Folksonomies, Tags, and User-Created Metadata: A Truly Emerging Technological Trend* Seminar on Emerging Technological Trends in Libraries. Reynolds Library, Wake Forest University. Retrieved from <http://laurenpressley.com/papers/folksonomies.pdf>
- Rothenberg, J. (2008). Interoperability as a semantic cross-cutting concern *Interoperabiliteit: Eerlijk zullen we alles delen*. Den Haag.
- Scott, H. (2007). *The Temporal Integration of Connected Study into a Structured Life: A Grounded Theory*. (PhD), University of Portsmouth Retrieved from <http://docs.google.com/viewer?a=v&pid=sites&srcid=Z3JvdW5kZWROaGVvcnlnbmxbmUuY29tfHd3d3xneDoyODQ1OWIwNmJmYzRhNzVj>
- Shadbolt, N. (2010). *data.gov.uk - The Linked Data Revolution*. Paper presented at the Innovating Through Information Lecture Series, London School of Economics, London. <http://eprints.ecs.soton.ac.uk/18787/>
- Shirky, C. (2005). Ontology is Overrated: Categories, Links, and Tags. *Clay Shirky's Writings About the Internet*. http://www.shirky.com/writings/ontology_overrated.html
- Shirky, C. (2008). *Here comes everybody : the power of organizing without organizations*. London: Penguin Books.
- Shirky, C. (2010). *Cognitive surplus : creativity and generosity in a connected age*. London: Allen Lane.
- Smith, G. (2008). *Tagging : people-powered metadata for the social web*. Berkeley, CA: New Riders.
- Spiteri, L., Tarulli, L., & Graybeal, A. (2010). *The Public Library Catalogue as a Social Space: Transaction Log Analysis of User Interaction with Social Discovery Systems*. Paper presented at the International Conference on Dublin Core and Metadata Applications Pittsburgh, PA.
- Spiteri, L. F. (2012). Social discovery tools: extending the principle of user convenience. *Journal of Documentation*, 68(2), 206 - 217.
- St. Pierre, M., & LaPlant, W. P. (1998). Issues in Crosswalking Content Metadata Standards. http://www.niso.org/publications/white_papers/crosswalk/
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Thousand Oaks, CA: SAGE Publications.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Thousand Oaks, CA: SAGE Publications.
- Sturtz, D. N. (2004). Communal Categorization: The Folksonomy.
- Styles, R. (2009). Bringing FRBR Down to Earth. Retrieved from <http://dynamicorange.com/2009/11/11/bringing-frbr-down-to-earth/>
- Styles, R., Ayers, D., & Shabir, N. (2008). *Semantic MARC, MARC21 and the Semantic Web*. Paper presented at the Linked Data on the Web (LDOW2008), Beijing, China. <http://events.linkedata.org/ldow2008/papers/02-styles-ayers-semantic-marc.pdf>
- Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge, Mass London: MIT Press.
- Tapscott, D., & Williams, A. D. (2010). *Wikinomics: how mass collaboration changes everything* (Expanded ed.). London: Penguin Books.
- Tennant, R. (2002). MARC must die. *Library Journal*. <http://www.libraryjournal.com/article/CA250046.html>
- Toffler, A. (1970). *Future shock*. London: Bodley Head.
- Toffler, A. (1980). *The third wave*. New York: Bantam Books.
- Udell, J. (2004). Collaborative knowledge gardening With Flickr and del.icio.us, social networking goes beyond sharing contacts and connections. *InfoWorld*. <http://www.infoworld.com/d/developer-world/collaborative-knowledge-gardening-020>
- Urquhart, C., & Fernandez, W. (2006). *Grounded Theory Method: The Researcher asBlank Slate and Other Myths*. Paper presented at the 27th International Conference on Information Systems (ICIS), Milwaukee.

- Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the 'theory' back into grounded theory: guidelines for grounded theory studies in information systems. *Information Systems Journal*, 20(4), 357–381.
- Van House, N. A. (2005). Science and Technology Studies and Information Studies. *Annual Review of Information Science and Technology*, 38(1), 1-86.
- Van House, N. A., Bishop, A. P., & Battenfield, B. P. (2003). Introduction: Digital Libraries as Sociotechnical Systems. In A. P. Bishop, N. A. Van House & B. P. Battenfield (Eds.), *Digital Library Use*. Cambridge, MA: MIT Press.
- Vander Wal, T. (2007). Folksonomy: Coinage and Definition. <http://www.vanderwal.net/folksonomy.html>
- Veltman, K. H. (2001). Syntactic and semantic interoperability: New approaches to knowledge and the semantic web. *New Review of Information Networking*, 7(1), 159-183.
- Virkus, S., Alemu, G. A., Demissie, T. A., Kokollari, B. J., Estrada, L. M., & Yadav, D. (2009). Integration of digital libraries and virtual learning environments: a literature review. *New Library World*, 110(3/4), 136-150.
- W3C. (2004a). OWL Web Ontology Language Overview. from <http://www.w3.org/TR/owl-features/>
- W3C. (2004b). RDF Primer. *W3C Recommendation*. <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/#basicconcepts>
- W3C. (2011). Library Linked Data Incubator Group Final Report: W3C.
- Wallis, R. (2011a). Library of Congress To Boldly Voyage To Linked Data Worlds. Retrieved from <http://consulting.talis.com/2011/11/library-of-congress-to-boldly-voyage-to-linked-data-worlds/>
- Wallis, R. (2011b). Will Europe's National Libraries Open Data In An Open Way? <http://dataliberate.com/2011/09/will-europes-national-libraries-open-data-in-an-open-way/>
- Weibel, S. (1999). The State of the Dublin Core Metadata Initiative: April 1999. *D-Lib Magazine*, 5(4).
- Weibel, S. L., & Koch, T. (2000). The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions. *D-Lib Magazine*, 6(12).
- Weinberger, D. (2005). Tagging and Why It Matters. <http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/07-WhyTaggingMatters.pdf>
- Weinberger, D. (2007). *Everything is Miscellaneous*. New York: Times books.
- Weinberger, D. (2012). *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room*. New York: Basic Books.
- Welsh, A., & Batley, S. (2012). *Practical cataloguing: AACR, RDA and MARC 21*. London: Facet.
- Welsh, E. (2002). Dealing with Data: Using NVivo in the Qualitative Data Analysis Process. *Forum: Qualitative Social Research*, 3(2).
- West, J. (2007). Subject Headings 2.0: Folksonomies and Tags. *Library Media Connection*, 25(7), 58-59.
- Williams, H. (2010). Linked Data and libraries. *Catalogue & Index* (160), 2-5.
- Wilson, N. (2010). *Linked Data Prototyping at the British Library*. Paper presented at the Talis Linked Data and Libraries event, British Library, London. <http://blogs.talis.com/nodalities/neil-wilson-lightning-talk-%E2%80%93-linked-data-and-libraries-2010>
- Wright, A. (2007). *Glut: Mastering Information Through the Ages*. Ithaca: Corenell University Press.
- Zeng, M. L., & Chan, L. M. (2006). Metadata Interoperability and Standardization – A Study of Methodology Part II: Achieving Interoperability at the Record and Repository Levels. *D-Lib Magazine*, 12(6).
- Zeng, M. L., & Qin, J. (2008). *Metadata*. London: Facet.

Appendices

Appendix-1: Ethical Approval Letters

Appendix-1.1: Ethical Approval of Study One (Proof-of-Concept Study)

Dr Steve C Hand
Head of School



Mr Getaneh Alemu
Creative Technologies
Eldon Building

8 June 2011

School of Creative Technologies
University of Portsmouth
Eldon Building
Winston Churchill Avenue
Portsmouth PO1 2DJ
United Kingdom

T: +44 (0)23 9284 5461
F: +44 (0)23 9284 5486
E: hod.ct@port.ac.uk
W: www.port.ac.uk/ct

Dear Getaneh,

I am pleased to inform you that the CCI Faculty Ethics Committee, based on the information you have provided in your initial application and your additional responses to our questions, has given your application for the study entitled '*Semantic metadata interoperability*', a favourable opinion.

This opinion has been given for this study only, and any changes in the conditions of the study may require you to re-apply for ethical review.

Although the Committee has given a favourable opinion, the final responsibility for the ethical conduct of this work lies, as always, with the researcher(s).

Please note that the Committee reserves the right to re-review this application should any concerns be raised about it in the future.

Your ethical review number is FO:01/11-0046

If you have any questions about this, please let me know.

Yours sincerely

A handwritten signature in purple ink, appearing to read 'Wendy Powell'.

Wendy Powell
(Chair, CCI FEC)

www.port.ac.uk

Appendix-1.2: Ethical Approval Letter for Study Two (Study with LIS Professionals: Researchers, Lecturers and Metadata Consultants)

Dr Steve C Hand
Head of School



School of
Creative
Technologies

School of Creative Technologies
University of Portsmouth
Eldon Building
Winston Churchill Avenue
Portsmouth PO1 2DJ
United Kingdom

T: +44 (0)23 9284 5461
F: +44 (0)23 9284 5486
E: hod.ct@port.ac.uk
W: www.port.ac.uk/ct

10th January 2012

Dear Getaneh

I am pleased to inform you that the CCI Faculty Ethics Committee, based on the information you have provided in your initial application and your additional responses to our questions, has given your application for the study entitled 'A Conceptual Framework for mixed- metadata approaches' (application date 19/12/2011), a favourable opinion.

This opinion has been given for this study only, and any changes in the conditions of the study may require you to re-apply for ethical review.

Although the Committee has given a favourable opinion, the final responsibility for the ethical conduct of this work lies, as always, with the researcher(s).

Please note that the Committee reserves the right to re-review this application should any concerns be raised about it in the future.

Your ethical review number is FO:01/12-0054.

If you have any questions about this, please let me know.

A handwritten signature in purple ink, appearing to read 'Wendy Powell', with a long, sweeping horizontal line extending to the right.

Wendy Powell
(Chair, CCI FEC)

www.port.ac.uk

Appendix-1.3: Ethical Approval Letter for Study Three (Study with Library Users)

Dr Steve C Hand
Head of School



School of Creative Technologies
University of Portsmouth
Eldon Building
Winston Churchill Avenue
Portsmouth PO1 2DJ
United Kingdom

T: +44 (0)23 9284 5461
F: +44 (0)23 9284 5486
E: hod.ct@port.ac.uk
W: www.port.ac.uk/ct

5th November 2012

Dear Getaneh,

I am pleased to inform you that the CCI Faculty Ethics Committee, based on the information you have provided in your initial application and your additional responses to our questions, has given your application for the study entitled 'A Conceptual Framework for the Inclusion of Socially-Constructed Metadata in Library Resource Description' (application date 11/10/2012) a favourable opinion.

This opinion has been given for this study only, and any changes in the conditions of the study may require you to re-apply for ethical review.

Although the Committee has given a favourable opinion, the final responsibility for the ethical conduct of this work lies, as always, with the researcher(s).

Please note that the Committee reserves the right to re-review this application should any concerns be raised about it in the future.

Your ethical review number is FO:11/12-0068

If you have any questions about this, please let me know.

A handwritten signature in blue ink, appearing to read 'W Powell', with a long horizontal stroke extending to the right.

Wendy Powell
(Chair, CCI FEC)

cc Brett Stevens

www.port.ac.uk

Appendix-2: Interview Questions

Appendix-2.1: Preliminary Study on Perspectives of LIS Academics and Post Graduates on Standards-based and Socially-Constructed Metadata Approaches (Study One)

This group of interviewees include library and information science academics and post graduate students. It is expected that the conversations held with this group of experts will provide a diverse mix of professional viewpoints on information organisation, metadata standards and metadata interoperability issues.

Notes to interviewer: Hand out printed copy of objective of the PhD research. Upon finishing the interview, thank participants for their participation, highlight on the importance of their participation and reassure them about the confidentiality of the information they provide. Note down the *name and contact details (preferably email address, place, date and time of interview) of each interviewee. Ask for consent for audio recording.*

Questions

1. Taking in to account your own personal and professional perspectives, how do you evaluate between browse-based navigation versus search-based navigation in electronic information services? *(This question is expected to answer: what is the experience of LIS academics and post graduates in using metadata to access information from websites, digital libraries and repositories? Answers to this question will help to gather participant views on their preferences between searches versus browse. Users will also be asked, to describe one or more of their typical search or browse strategies. This will help to conceptualise what makes them to decide to prefer one navigation technique over another; for instance if taxonomies (labels), search terms, etc. have an effect on users' experiences. This will also help to look into the user behaviour patterns between search and browse).*
2. From your experience, how is the Online Library Catalogue (OPAC) situated in the digital information landscape? *(The answer to this question will help to conceptualise how the participant's view of the current status of the library's OPAC. The answers may help to indicate how library metadata in general and the OPAC interface in particular is viewed by LIS academics and post graduates in relation to its efficacy in helping users find and discover information objects. This will inform if there is a pattern in shift of user behaviour in the use of OPACs. If there is such change in pattern, to determine the underlying causes and identify what metadata approaches should be adopted in the future?)*
3. Tell me how you think standardised and centrally controlled subject headings (taxonomy, thesauri, Topic Maps) and classification systems (DDC, LC) are optimised for use in today's information landscape? *(The assumption is that as LIS postgraduates and academics may be familiar to these standards in their studies and researches, their perspectives on how these standards-based metadata approaches are addressing their professional and personal information needs. What are the strengths and limitations of these standards?)*

4. How do you think such issues of synonyms and homonyms or semantic heterogeneity in library metadata should be addressed? *(This will help to identify the participant's perspectives on how current standards address the issue of semantics and terminological differences (cultural diversity). How do you think should current metadata standards reflect changes in the meaning of values in their metadata?)*
5. Which metadata standards have you used in your professional works? *(This helps to identify participants' experience and exposure to the various metadata standards.)*
6. What do you think are the problems of existing metadata standards to support interoperable access between disparate digital repositories?
7. Currently there is a diversity of metadata standards in the library and archives community, some argue that standardisation towards a single standard is the solution? Some say local standards should be allowed to develop and interoperate to that of the international? *Who should be concerned about interoperability? Who should pay for it? How do you think an institution can mix-and-match metadata elements from disparate standards? How can semantics be addressed? Language and cultural differences? Not-invented-here syndrome?*
8. What is your view on social tagging and folksonomies (bottom-up metadata)?
9. Ask them if they have used one. What are their issues? How did they resolve those issues?
10. What is the Semantic Web for you?
11. As LIS professionals, what is their understanding of the phrase? How does it affect them?
12. Do you have anything to add? Is there anything you would like to ask?

Thank you very much for your time

Appendix-2.2: A Matrix of Semi-Structured Interview Questions and Constructs (Perspectives of LIS Professionals on Standards-based and Socially-Constructed Metadata Approaches) (Study Two)

[illegible]

Appendix-2.3: Semi-structured Interview with Library Users (Study Three)

Notes to Interviewer

Following a Constructivist Grounded Theory Method implies that both the interviewer and interviewee are actively engaged in conversations. Intensive interviews are conducted with open-ended questions and lengthy engaging discussions. According to Charmaz (2006, p.25) “the in-depth nature of an intensive interview fosters eliciting each participant’s interpretation of his/her experience”. Consent of interviewees need to be sought for time, place and voice recording arrangements. As Charmaz advises intensive interviews should be contextual and negotiated. It is therefore important to ask precise and short lists of questions and keep the interview as informal and conversational as possible. Intensive interviews allow the interviewer to express views and reflect on earlier points and share experiences (Charmaz, 2006, pp.26-27). A concise summary of the objectives of the research and the confidentiality of the information obtained from the interviewees need to be stated before the start of the interview.

Interview Library Users

With the aim of developing a conceptual metadata framework, this study explores the perspectives of library users their use of current library collections using the library’s Online Public Access Catalogue and also their views on the roles of user-driven resource description including user reviews, tags, ratings and recommendations will be explored. A purposive (opportunistic sampling) will be used to identify and select participants. Pursuant to the Grounded Theory Method’s concept of theoretical saturation, the size of the interviewees cannot be predetermined at first; instead the process will be evolving based on the analysis of the first sets of interviews. The result of this data collection is expected to provide information on the experience of library users in searching/browsing information objects such as books, and articles. The study explores the perspectives of users on Online Public Access Catalogues. Users will be asked of their experiences in using online databases and services and the discussion will be contextualised to their experiences in using library catalogues. From the interview discussions, it is expected that relevant concepts will emerge to inform this research in relation to the preferences and requirements of library users. The interviews will be audio-recorded, transcribed. NVivo 9 qualitative data analysis software will be used to code and analyse it. The data will help to saturate existing categories that emerged from the first two studies and may also result in new categories which will be integrated into the analysis and discussion.

Notes to interviewer: Hand out printed copy of objective of the PhD research. Upon finishing the interview, thank participants for their participation, highlight on the importance of their participation and reassure them about the confidentiality of the information they provide.

Note down: *The name of interviewee, contact details (preferably email address, place, date and time of interview. Ask for audio recording consent even if that means it was asked in the consent form.*

Interview Questions (Study Three)

1	Taking into consideration your own experience in finding and discovering books, articles, and related information objects for your teaching/research/studies? Would you please describe how you go about finding/discovering/identifying these resources? (Constructs: <i>identifying, finding, discovering information objects</i>)
2	How do you use the library? Do you physically go or use it electronically? How do you find books etc. through serendipitous discovery than known-book search? (Constructs: <i>Use of library, use of library resources, known item searching, discovering serendipitously</i>)
3	Do you use the library OPAC/search engine to find books in the library? (Constructs: <i>Use of library catalogue, searching the OPAC, navigating the OPAC, OPAC interactivity</i>)
4	Is the book information displayed on library search results/OPAC searches adequate? Do you prefer more detailed information about a book such as Table of Contents, book jackets, subject keywords, or just basic metadata (information) such as author, title, date of publication? (Constructs: <i>OPAC interface/display, OPAC display richness</i>)
5	How do you describe the current status of your library OPAC (strengths and weaknesses)? (Constructs: <i>OPAC strengths, OPAC weaknesses</i>)
6	If you can change the OPAC, what aspects of it would you like to change? (Constructs: <i>Future of OPAC</i>)
7	Do you think users should be involved in the description (tagging books and articles of their preference for latter retrieval, writing comments and reviews about books they have consulted/read, sharing book recommendations among peers, rating the relevancy of books for a particular study/research, etc.)? What implications do you think such user-generated metadata may have on the library's services? (Constructs: <i>tagging, ratings, reviews, recommendations, implications of user-generated metadata on library services</i>)
8	How do you think, if you are familiar with it, Amazon.com-like recommendation services would be relevant in libraries? (Constructs: <i>relevancy, benefits, and limitations of user and system recommendations in the library context</i>)
9	If the library implements, social tagging/recommendations/ratings/reviews of books, would you be willing to contribute metadata? (Constructs: <i>User participation, interest, motivation</i>)
10	Should user-contributed/created metadata such as reviews, tags, ratings, etc. be monitored and controlled by librarians? (Constructs: <i>Metadata quality, metadata vetting/policing, idiosyncratic (personal) metadata, user involvement in social metadata quality control</i>)
11	What do you think motivates users to contribute metadata? (Constructs: <i>User motivation, personal motivation, altruistic motivation</i>)
12	What benefits do you think user-created metadata would have to improve your information finding/discovering experience? (Constructs: <i>benefits, implications on information findability/discoverability, implications on personal study/research</i>)
13	What intellectual property (metadata ownership) and privacy issues do you see in user-created metadata? If any, how should it be resolved? (Constructs: <i>Metadata ownership (copyright) issues, privacy issues</i>)

14	Do you think linking the data displayed in the OPAC/library search interface to external information services is useful or unnecessary cluttering of the library website? (Constructs: <i>Linking between databases, linking to external information services, metadata richness</i>)
15	Do you have anything to add? Is there anything you would like to ask?

Thank you very much for your time!

Appendix-3: Glossary

A priori Metadata: A priori metadata can be defined as the metadata created and affixed to an information object before “publication”, i.e. before release of information to users. It can thus be considered as before-the-fact metadata, as the description of the information object is made before its use.

Category: In grounded theory’s Focused Coding stage of data analysis, Open Codes from the Open Coding stage are further organised in to a nested hierarchy called Categories.

Coding: In Grounded Theory Method, the process of coding refers to the technique of labelling pieces of selected interview text (or other qualitative data) with terms. The purpose of coding is identification of concepts from data. There are three stages of coding: open, focused and theoretical coding.

Core Categories: During grounded theory’s Theoretical Coding, the Categories from the Focused Coding are compared and integrated to form what is referred as Core categories. Core categories are emergent and grounded in empirical data analysis.

Focused Coding: In Grounded Theory Method, focused coding involves bringing similar codes into a nested category of codes. Focused Coding is an abstraction of the codes generated during the Open Coding stage. This stage of data analysis results in comparing, linking, refining and abstracting open codes into a hierarchy of codes called categories.

Grounded Theory Method: It is an inductive research method where it aims to derive concepts and theory from the data. It is an iterative and evolutionary research methodology where simultaneous data collection and analysis takes place.

Linked Data: As the name indicates, Linked Data is a data model that identifies, describes, links and relates structured data elements, analogous to the way relational database systems function, albeit the fact that Linked Data is aimed to operate at a web scale.

Metadata Actors: Also called metadata creators. Four groups of metadata creators can be identified: librarians, authors, users and machines.

Metadata Filtering: Filtering is the process of sorting, sifting and presenting metadata in discovery interfaces (OPACs). A priori filtering can be done by librarians, whereas post-hoc filtering can be done by users themselves. The principle of filtering states that metadata can be presented to users through a plethora of ways including using keywords, their profiles, interest, and other metadata facets.

Metadata: Is often defined as data about data. It is that data that describes an information object, for example title, author, year of publication, subject keywords of a particular book.

Objectivist View: It is a philosophical view which assumes that an objective reality exists independent of the observer. It is contrasted with social constructivism.

Open Codes: During the Open Coding stage, segments of interview transcript are coded with labels. The resultant labels are referred in this thesis as Open Codes.

Post-hoc metadata: In contrast to a prior metadata, post-hoc metadata is created by users after the publication of information objects.

Resource Description Framework: RDF is a data model to describe any concept or object (physical and abstract) using simple Subject-Predicate-Object (also called triple) statements.

RDF Schema (RDFS): In order for the RDF model to function, it requires defined vocabularies, thus RDF schema provides specifies vocabularies such as Class, SubClassOf, Domain, Range, Label and Comment.

Serendipity: Serendipitous discovery of information objects is related to the possibility of accidental but fortune discovery of information objects.

Social Constructivism: Social constructivism is philosophical view point which posits the possibility of the existence of multiple realities and interpretations. It is discussed in contrast with the objectivist view.

Socially-Constructed Metadata: The phrase socially-constructed metadata refers the practise of users creating metadata (i.e. user-generated metadata). Socially-constructed metadata includes, for example, tagging, user reviews, recommendations, and ratings.

Standards-based Metadata: Agreed upon principles, rules and procedures for creating and maintaining metadata, often designed by international (top-down) standards-setting agencies so as to be used by experts in creating metadata for information objects. Metadata standards include MARC, DC, FRBR, and RDA.

Theoretical sampling: In grounded theory data analysis, theoretical sampling refers to the process of selecting categories that are considered pertinent for developing core categories. Theoretical sampling informs the nature and source of data to be collected to further develop the categories.

Theoretical Saturation: In the context of the Grounded Theory Method, theoretical saturation is the stage at which the researcher makes a decision because the core categories that have emerged from the research process are saturated (developed) with adequate data to the extent that the incorporation of new data provides no additional insight.

Uniform Resource Identifier: A Unique Resource Identifier is a unique web-based global identification assigned to information resources and metadata. One of the defining features of the RDF model is its ability to identify resources and metadata attributes (relations) uniquely and globally using URIs.

Appendix-4: FORM UPR16 (Research Ethics Review Checklist)

Please complete and return the form to Research Section, Quality Management Division, Academic Registry, University House, with your thesis, prior to examination



Postgraduate Research Student (PGRS) Information		Student ID:	479615
Candidate Name:	Getaneh Alemu		
Department:	Creative Technologies	First Supervisor:	Dr Brett Stevens
Start Date: (or progression date for Prof Doc students)		October 2010	
Study Mode and Route:	Part-time <input type="checkbox"/> Full-time <input checked="" type="checkbox"/>	MPhil <input type="checkbox"/> MD <input type="checkbox"/> PhD <input checked="" type="checkbox"/>	Integrated Doctorate (NewRoute) <input type="checkbox"/> Prof Doc (PD) <input type="checkbox"/>
Title of Thesis:	A Theory of Digital Library Metadata: <i>The emergence of Enriching and Filtering</i>		
Thesis Word Count: (excluding ancillary data)	80,000		

If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study

Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).

UKRIO Finished Research Checklist:

(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: <http://www.ukrio.org/what-we-do/code-of-practice-for-research/>)

a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	YES
b) Have all contributions to knowledge been acknowledged?	YES
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	YES
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	YES
e) Does your research comply with all legal, ethical, and contractual requirements?	YES

*Delete as appropriate

Candidate Statement:

I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)

Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):

1. FO:01/11-0046
2. FO:01/12-0054
3. FO:11/12-0068

Signed: Getaneh Alemu

(Student)

11-03-2014